

Science and the theory of history

General excitement about the argument by Francis Fukuyama that recent setbacks for totalitarian states mark the end of history should be moderated by awareness of the weaknesses of his interesting case.

HISTORIANS, the world knows, are maddening people. If asked whether history has a pattern, they will usually assert that the mere question is offensive; history is but the record of the chaotic events that come about because people's hopes and ambitions are invariably modified by external circumstances, often the hopes and ambitions of other people. An historian's job, the argument goes, is not to find a pattern, let alone a 'law', but to ensure by research that the record is accessible and intelligible. But not all historians are that purist. There are, for example, Marxist historians who look preferentially for patterns they believe reveal the perpetual struggle between the proletariat and its masters. Others — usually considered second-rate historians — hold to the idea of progress, or the notion that the improvement of the human condition in recent centuries can, with ingenuity, be continued into the future. Natural scientists often fall into that camp.

Francis Fukuyama is not an historian in the strict sense, and is more a political scientist than a sociologist. His affiliation is with the Rand Corporation of Santa Monica, originally a defence consultancy, which now takes on civilian business. Fukuyama has recently startled people in the United States and elsewhere with his book *The End of History and the Last Man* (Free Press, New York, and Hamish Hamilton, London), stacks of which can now be found in most bookshops. One reason for the general excitement is that the book is beautifully written; it is not often, these days, that one can buy 400 pages of literate text for the price of a cab fare to a metropolitan city airport. But the book has caught the headlines because it is a hymn to the virtue and success of the system of government called liberal democracy. The most superficial reading is that the collapse of Communist governments in the Soviet Union and Eastern Europe in the past three years marks the inexorable triumph of liberal democracy over totalitarianism.

It will be a misfortune if that is all Fukuyama is remembered for. In particular, his argument entails suppositions and assertions about the historical role of science and technology that deserve more attention than they have so far been given. A few preliminary definitions are necessary. A liberal state is one that respects individual rights, as in the ownership of property, but is not on that account democratic (witness the counter-example of Singapore). A democracy is a state in which the people choose their government on the basis of

universal suffrage, and need not necessarily be liberal (witness Iran, which holds regular elections, but is religiously intolerant). Fukuyama argues empirically that states that are both liberal and democratic are on the increase. So far, so what?

The essence of Fukuyama's case is that it must be so. But to assert that illiberal and undemocratic states are unstable relative to liberal democracies is to put forward a theory of history, which is why conventional historians regard Fukuyama as an upstart. But the title gives the show away. The "end of history" refers to nineteenth-century Hegel's belief that it should be possible to write a "Universal History", a theory of everything that has ever happened; Fukuyama's vision of the end point is that in which almost all states are liberal democracies. And the "last man" refers to Nietzsche's dour vision of the people inhabiting that homogenized world, where the avoidance of controversy is a selective advantage. Fukuyama's question (not answered in the text) is whether we (or our successors) will be happy that way.

Natural science enters this theory of history most simply as a kind of ratchet; discoveries once made are not easily forgotten, which means that for at least the past 400 years, history must have been directional. But Fukuyama also argues that technology and its flair for producing goods that people want also creates in liberal societies the need for democracy. Yet if technology is necessary, it is not sufficient to bring about that transformation. It is also necessary that people should fight (and risk their lives) in the democratic cause. And people do it all the time. Fukuyama mentions Yeltsin standing on a tank outside the Russian Parliament last August.

This is not, of course, the first time that people have argued that science and its application are important instruments for the improvement of society. Once Europeans had shaken off the classical myth of the Golden Age (the earlier time when everything was supposedly better), the doctrine of improvement came to take its place. Robert Nisbet (in *The History of the Idea of Progress*, Basic Books, New York, and Heinemann, London; 1979) has catalogued a rich stream of literature beginning in the sixteenth century arguing that both scientific knowledge and its application make society more civil. Francis Bacon is only one of the best known among these early exponents of Fukuyama's case, but Adam Smith was even more influential.



Fukuyama's argument is the more arresting because the evidence on which it mostly relies is contemporary history, as recorded by the newspapers we read. But especially for that reason, it is curiously incomplete. This theory of history, for example, relies for generalizations about human behaviour on the observations of people like Hegel and Nietzsche rather than on Freud or, for that matter, B. F. Skinner. More seriously, the general conclusion that the problems that lie ahead consist largely of finding interesting occupations for endless replicas of the last man begs several teasing questions, from global warming and the consequences of rapid population growth to the difficulty of reconciling Muslim states to the notion that liberal democracy is indeed the end of history. The trouble is that all these problems are, in their different ways, potential threats to liberal democracy. It is a pity that Fukuyama has ducked them. □

Mystifying manifestos

The two major British political parties seem to have turned their backs on basic research.

ALTHOUGH nobody quite knows why political parties publish political manifestos, election-struck Britain has been showered with them in the past few days. The most striking is that published as a newspaper advertisement (at an estimated cost of £200,000) which includes simulations of formulae that might well have been taken from a textbook of elementary quantum mechanics. That party consists of people persuaded that transcendental meditation solves personal problems and may thus solve the world's; it is not expected to do well in the general election now fixed for 9 April.

The three serious parties, the government (Conservative) party, the chief opposition (Labour) party and the Liberal Democratic party, on the other hand, have put out glossy brochures stuffed with promises of what they would do if elected to form a government. To tell from these documents, British science will do no better in the future than in the past, whichever party is elected.

The Conservative and Labour manifestos do not mention research as an objective in itself. Literally, in neither does research rate a mention. Only the Liberal Democrats refer to research as such, but then only cursorily, by comparing the proportion of Gross Domestic Product (GDP) spent on basic research in France and Germany with the shrinking proportion (now 0.28 per cent) spent in Britain. The Liberal Democrats go on to say that they would restore the proportion of public spending on civil science to 0.35 per cent of GDP, but even that is a smaller proportion than richer Germany spends.

That is a mystifying state of affairs. Even the rawest recruit to one of the major parties' manifesto-writing teams could have turned out, if asked, an uplifting sentence or two about the folly of wasting the seed-corn, or about the reasons why discoveries are the well-springs of industrial innovation and thus are the foundations of

national prosperity. With care and a little supervision, the sentences could even have been made promise-free. Yet either these uplifting sentences were never written or, alternatively, were not selected by those who merged the word-processed files that went to make the two chief parties' manifestos. And why should that be?

On the evidence of last week's brochures, the British scientific community must face the stark truth that there is now bipartisan agreement that basic research is either unimportant or unmentionable — or both. Last week's manifestos point to the last interpretation. While failing to mention basic research, each major party promises to spend much of its energy on the provision of vocational training. The government would build on existing Technical Education Councils and on its plans for vocational qualifications for school-leavers; Labour would make a "coherent" package of such schemes. If the premise is that the British labour-force is not sufficiently educated, few will quarrel. But is it intended that radical innovation should sit on the back-burner until that deficiency has been remedied? □

Imaginative biology

The Human Frontier Science Programme, inspired by Japan, deserves more general attention and support.

THE Human Frontier Science Programme is a remarkable venture, reflecting most of all the sheer daring of the Japanese view of how science and technology advance. Five years ago, when the notion of an international collaboration in the field was first canvassed by Japan, many in the West held that the prospectus was too vague to be taken seriously. Unwisely, many chose not to notice what was happening. But the truth is that there is now, at Strasbourg, the first organization to be a genuinely international grant-making research organization. People can apply for research funds, be sure that their applications are properly assessed by others in the field and may, with luck, be given funds to spend. That is something for the government of Japan and for Sir James Gowans, the first secretary general, to be pleased about.

The planned changing of the guard next year (see page 277) should be an opportunity for more fun and games. From the outset, the Japanese have looked to the Human Frontier Science Programme as a means of stimulating imaginative ways of tackling problems in biology. What, for example, do control engineers have to say about homeostasis? Can physical chemists so specify the conditions satisfied by self-assembling molecular aggregates in biology — synapses or muscle fibres, for example — that the natural functions of these systems would be better understood and artificial analogues developed for industrial use? Is it even time to embark on the construction of a mathematical model of a cell, no doubt the megaproject that will follow the human genome projects? The new partners at Strasbourg should not be afraid to let Japan push them in such directions. □

Japan stubs its toes on fifth-generation computer

- One-year extension to save face
- US collaboration called a sham

Tokyo & Washington

JAPAN'S once mighty fifth-generation computer project, which in the early 1980s sent Western governments scurrying to set up competing computer initiatives, has fallen far short of its mark.

The 10-year project was due to finish at the end of this month, but it has lagged so embarrassingly behind schedule that the Japanese government has given the project an extra year to inch closer to its target.

Launched in 1982 by the Ministry of International Trade and Industry (MITI), the fifth-generation project set out to develop a massive 'user friendly' parallel computer with 1,000 processors. Its software was based on logic rather than on classic structured programming. The Japanese initiative raised fears that Japan was about to take over the world's advanced computer market and sparked many competing projects in the West, such as the UK's Alvey project.

But, after ten years and an investment of more than \$400 million, all Japan has produced is a handful of parallel computers, the largest three of which have only a quarter of the target number of processors. And only a limited number of computer experts know how to operate them.

Despite its failings, the project is receiving an extra £3,600 million (nearly US\$30 million) from MITI in fiscal year 1992, and project researchers are rushing to assemble a 512-processor machine in time for a final international conference on the project in June.

The fifth-generation computer project is a classic example of how the rigid Japanese bureaucracy can foil a national project. By the mid-1980s, it was clear that other approaches to parallel computing not based on traditional artificial intelligence techniques, such as neural networks or the massively parallel machines created by Thinking Machines Inc. of the United States, looked more promising. But having told the Ministry of Finance that it would build a 1,000-processor machine, MITI had no choice but to continue towards that goal. MITI officials fully realize their failings, however, and in their next 'sixth generation' computer project they intend to set more flexible goals (see next page).

One ray of hope for the fifth-generation computer came in 1990. Some of the researchers at the project's Institute of New Generation Computer Technology (ICOT) in Tokyo established a high-capacity computer link with the US Argonne National Laboratory to try and use one of the prototype fifth-generation computers in ICOT to solve biological problems associated with the human genome project (*Nature* 345, 466; 1990).

But that collaboration went disastrously awry (see sidebar), and similar collabora-

tion with the US National Institutes of Health (NIH) and the Lawrence Berkeley National Laboratory (LBL) are essentially moribund. US researchers found the fifth-generation computers to be slow, cryptic and filled with bugs. And the documentation that came with them was written with Kanji characters, which were never translated into English.

As a result, the fifth generation machines in the United States are now "essentially doorstops" that are used mostly for electronic mail, says one researcher. At NIH, "we're not using their hardware or their software," says George Michaels, a NIH geneticist. US researchers say that the Japanese were more interested in showing the world that the machines were being used by NIH and US national laboratories than in any real collaborations. "The reason the machines came here was essentially PR," Michaels says.

(Continued on page 274)

Collaboration vetoed

In January 1991, with the United States preparing for war in the Middle East and anti-Japanese sentiment running high, officials at the US Department of Energy (DOE) shut down one of the three US collaborations with the Japanese fifth-generation computer project in fear of a backlash from Congress.

The official reason was that the US team at DOE's Argonne National Laboratory had not obtained proper approval before visiting Japanese laboratories and accepting two computers from them. But, in fact, they acted after reading an article about the collaboration in *Supercomputing Review*. Being seen as sharing technology with such a fierce rival, DOE officials told scientists, could cause problems for the agency when it came time to defend its budget before Congress.

In October, the principal US collaborator in the project, Ross Overbeek, resigned in protest after a long war of memoranda. Overbeek declined to discuss the case, but others say that he objected to the DOE claim that the collaboration might lead to "inappropriate technology transfer to the Japanese."

In fact, no technology was transferred — but not because the Argonne team had refused to do so. The Japanese themselves had shown little interest in the software the Argonne researchers had developed.

Ironically, the collaboration had effectively ended before the controversy began. Similar collaborations at the National Institutes of Health (NIH) and the Lawrence Berkeley National Laboratory were not affected by the DOE decision, mostly because the officials who ran those programmes were less concerned about the politics of Japanese collaboration.

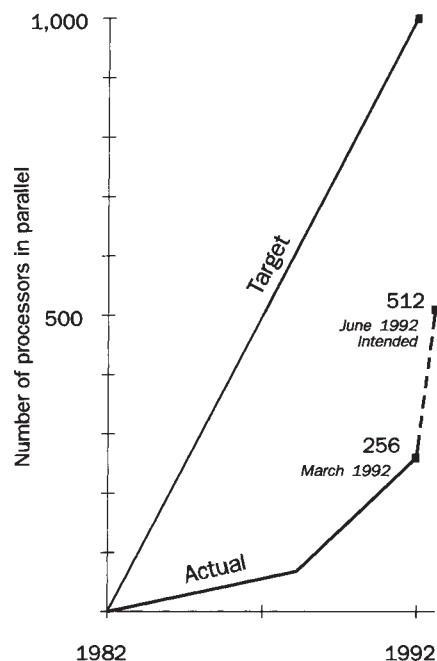
First announced in June 1990, the Argonne collaboration was intended to develop software collaboratively that could use the logic-based techniques of the fifth generation computer project to analyse genome data (see *Nature* 345, 467; 1990). But like other US-Japanese collaborations in the project (see adjacent story), it quickly ran into technical problems.

"On the surface, it looked like Washington put pressure on the collaboration to stop," says Rick Stevens, director of Argonne's mathematics and computer science division. "But from our standpoint, it wasn't going anywhere, anyway."

Indeed, DOE officials say, that lack of progress was part of their concern. Collaboration should be "symmetrical", says David Nelson, DOE's director of scientific computing. "Just because someone gives you a free machine," he says, "doesn't make it a collaboration."

Outside observers have another perspective. "It was an amazing overreaction" in a time of political tension, says George Michaels, a NIH genetics researcher. C.A.

Short of the mark



(Continued from page 273)

But the project has nevertheless left a legacy in the US genetics community, researchers say. Genetic manipulation software being developed both at NIH and at LBL uses many of the same logic programming techniques as the original fifth-generation software, if none of its actual components. And the project led US genetics researchers to talk among themselves about new computer techniques, says Cassandra Smith, who heads the LBL team. Those collaborations, at least, are still bearing fruit.

ICOT has also tried to link up with biologists at Kyoto University. As in the case of Argonne and NIH, the institute donated personal sequential inference machines to the university that could, in theory, be linked to the mother computer in Tokyo. But because ICOT is not a university and is not part of MITI, its scientists have not been able to establish a computer link.

Nevertheless, MITI officials see the genome project as one possible way to keep at least a part of the fifth-generation project alive after the end of the current fiscal year. And the biochemical industry division of the ministry will soon form a committee to coordinate this and other genome-related research activities supported by MITI (*Nature* **356**, 181; 1992).

However, companies which have been participating in the fifth-generation project have no interest in seeing it continue. They are tired of having some of their best researchers tied up in the project, particularly now that industrial research budgets are being cut. ICOT has 90 researchers, nearly all of them drawn from industry, and another 200 researchers are working for the project at their companies.

Some young ICOT researchers are very keen for the project to continue so that they can develop and test out software on it. But ICOT's director, Kazuhiro Fuchi, thinks it will be very difficult for the project to continue beyond March 1993.

**David Swinbanks &
Christopher Anderson**

Classification Catch-22

Washington

US DEFENCE officials have accused an outspoken critic of the Patriot missile and the "Star Wars" missile defence system of publishing secrets. But they are unable to prosecute him because he will not allow them to tell him what those secrets are.

The critic, Theodore Postol, a Massachusetts Institute of Technology (MIT) physicist, says that he used only unclassified data in his calculations for an article that claims that the Patriot missile was "an almost total failure" in the Gulf War. But if he lets defence officials identify what they believe is classified information in the article, he says, it will become by definition secret and he will not be allowed to talk about it, even if it was based on unclassified data.

This Catch-22 may have put Postol in the public eye for the moment as the latest in a list of researchers who have run aground in this still uncharted corner of the classification rules. The problem is a concept known as 'compilation'. Simply put, when nominally unclassified data is assembled in such a way that the end result is more secret than the sum of its parts, it can become too secret to publish.

In 1979, *The Progressive* magazine got into trouble with an article that described how to build a hydrogen bomb. Culled from unclassified interviews with nuclear scientists and publicly available information, the article triggered a lengthy legal battle before it was finally allowed to be published.

A few years later, Bruce Blair, an analyst at the Brookings Institute in Washington, DC, found that he had crossed the line in compiling an unclassified report about the electromagnetic pulses that follow nuclear explosions. Even Blair himself was not allowed to keep a copy of this suddenly secret report. Since then, even Stansfield Turner, a retired admiral and former director of the Central Intelligence Agency, ran afoul of officials who sought

to remove details from his 1985 book, *Secrecy and Democracy*. They told him that, while the information was not secret, it had been compiled with the help of classified knowledge.

In Postol's case, there are actually two issues. The first is whether there is, indeed, anything secret about his article, which appeared earlier this year in *International Security*. He says no, and he has invited defence officials to check his claim by examining all 100 unclassified references. Postol says that he obtained other figures in the article by doing simple calculations based on publicly available information; to determine the top speed of a Patriot missile, for example, he did a mass distribution calculation based on an unclassified photograph.

But the other issue is trickier. Before coming to MIT, Postol held a top-level classified position with the Navy, where he evaluated advanced weapons. Since then, he has retained his security clearance and recently sat in on two classified talks on issues relating to the Patriot. He says neither of the talks provided him with any information for the article.

He has offered to show the classification officials where and how he obtained each figure in his article. But he says they do not want to bother, and they have told him not to discuss the article until the issue is resolved. Last week, he told the Government Operations Committee of the US House of Representatives that such abuses of the classification system censor free speech and "pose one of the most serious and overriding threats to democracy and its institutions".

Postol's case rests with Congress and the Defense Department. But critics of the current classification system say there are many similar cases in which reports and papers based on unclassified disappear after they are deemed too sensitive. "It's an all-too-common extension of an already aggressive classification policy," says Steven Aftergood, a security expert with the Federation of American Scientists.

Turner calls it "a way of improperly classifying material". Avoiding a 'secret' stamp on sensitive articles is not easy, he says, especially if a researcher has a security clearance. "You have to establish that an ordinary person without special skills couldn't have done [the work]" it says. "While there's no law that says you need to prove your innocence, that in fact is what you have to do."

One solution to the problem is to avoid issues that embarrass the government. Failing that, says Roy Woodruff of the Los Alamos National Laboratory, "use good research judgement and dig in your heels when they complain."

Christopher Anderson

Talking about a new generation

JAPAN'S Ministry of International Trade and Industry (MITI) is determined to avoid some of the pitfalls of the fifth-generation computer project (see previous page) as it begins a follow-up computer project of comparable scale and duration.

Rather than setting its sights on one specific type of hardware and software, the ministry is planning a multi-pronged approach that will include development of neural network and optical computing systems as well as a general-purpose massively parallel computer. In another departure from the earlier project, MITI plans to farm out much of the work to companies, national research laboratories and universities both within and outside Japan.

MITI also seems determined to erase any connections with the preceding fifth-generation project. MITI has named it 'four-dimensional' (*yojigen*) computer for home consumption and the 'real-world computing program' in English.

Once again, however, MITI is setting itself a very ambitious and specific target. Its ten-year goal is to build a computer with one million processors, some thousand times the goal of the fifth-generation project.

D.S.

New company to target technology transfer

Paris

FRENCH research organizations are forming an independent company to work with industry to translate their discoveries into commercial products.

The new company, which will be launched in June, plans to offer government scientists full intellectual property rights and full royalties on licences from patents based on their research in exchange for the laboratory paying all costs for the transfer of the technology to the private sector. Laboratories that prefer less risk can pay a portion of the costs in exchange for a negotiated division of the royalties.

The company will take over some of the duties of the national agency charged with promoting innovation as well as promoting technology transfer. The agency, called ANVAR (Agence Nationale de Valorisation de la Recherche), decided last year to concentrate on promoting and financing innovation. That left a vacuum that the ministers of Industry, Dominique Strauss Kahn, and of Research and Technology, Hubert Derian, rushed to fill.

A study by Jean-Claude Derian, former science adviser at the French embassy in Washington, called for a company with a

budget of FF50 million (US\$8.5 million) that would both broker licences and provide venture capital for promising ideas. The structure agreed last month, however, will focus on patenting and licensing and have a budget of only FF5 million (\$850,000) and a staff of seven. Jean Remi-Gouze, deputy executive director of ANVAR, explains that "the Ministry of Finance has argued that the creation of a state-owned venture capital company is unnecessary".

The Centre National de la Recherche Scientifique (CNRS), the agency that funds most basic research in France, and ANVAR will each provide about one-third of the total, with smaller amounts coming from other government funding agencies and universities. Opting out for the moment are the Institut National de la Santé (INSERM), which funds biomedical research, and the Commissariat à l'énergie Atomique (CEA), which funds energy research, both of which have well-developed technology transfer offices.

The origins of the new company can be traced to 1982, when the French government broke ANVAR's monopoly on technology transfer by allowing laboratories greater freedom to exploit their inventions.

Since then, technology brokerage organizations have proliferated, but for the most part they have specialized in transferring their technologies to local enterprises. Few organizations have had the means to work with international companies, however, a major goal of the new company. Officials point out that the bulk of the revenues from the British Technology Group, for example, comes from foreign sources.

The new company will survive only by acquiring a sufficient number of patents. It is counting on most of ANVAR's current portfolio, which includes a large number of CNRS patents, and adding new patents at the rate of 125 a year. "If it doesn't achieve these targets", warns Gouze, "it will go under".

The almost unanimous backing of the French research establishment is expected to help the company to gain credibility on the international scene. But its chief executive officer, to be named soon, must also possess a rare combination of technical, entrepreneurial and administrative talents. Gouze says the government is looking for what the French call "un mouton à cinq pattes", or, translated loosely, something that does not exist.

David Bakewell

CHINESE REFORMS

Scientists emerge with favoured status

Hong Kong

CHINA's scientific community stands to be a major beneficiary of the new economic reforms advanced by Deng Xiao-ping, the country's 87-year-old leader. With bonuses for thousands of civilian scientists, fewer restrictions on travel, more joint ventures with outsiders and increased spending on research, the importance of technological progress is expected to be a major theme of the current annual session of the National People's Congress, which opened last week in Beijing.

The link between reform and science was made clear earlier this month at a meeting of the State Science and Technology Commission. Li Xu, vice-minister of the commission, declared that "establishing more share-holding companies and issuing bonds in [high-technology] districts... are effective ways to accumulate capital for money-starved research". This week's congress is expected to adopt a measure that would guarantee investment in science, require agriculture and industry to adopt the latest technologies, encourage international cooperation and protect the rights of scientists.

Some 27 high-technology development districts throughout China will receive the largest measure of freedom. These districts, the first of which was set up four

years ago in Beijing, are seen as the most promising way to bring the country's products to the international market.

But not every aspect of the plan is on such a grand scale. Individual scientists and research teams will be paid bonuses for outstanding achievements. The Ministry of Personnel intends to award an extra 100 yuan (US\$18.50) a month to "the most promising middle-aged and young scientists". That is a significant addition to the US\$100-120 paid monthly to the average faculty member.

Last year, more than 9,000 people described as "experts and scholars" received such extra allowances. That group is believed to include military researchers, doctors and technicians. The new program extends such bonuses for the first time to large numbers of scientists in the civilian sector. "We guarantee that outstanding enterprises and individuals will get high pay," says Li.

Deng began his campaign in January with a well-publicized visit to Guangdong, the province in southern China in the forefront of the country's modernization. The special economic zones at Shenzhen and Zhuhai have tapped into sizeable investments by high-technology companies in neighbouring Hong Kong and Macao.

A recent report by the science commis-

sion noted that "we encourage Chinese enterprises to set up joint ventures, research institutions and information service units in foreign countries". Speaking to the *China Daily*, Beijing's official English language newspaper, Li was quoted as saying that "the process of leaving and returning to the country will be greatly simplified. Scientists are now free to come and go."

China is also looking overseas for scientists. Reports from the meeting of the science commission suggest that the government wants to import technicians and scientists from the former Soviet Union, in particular from Russia, Ukraine, Byelorussia, Uzbekistan and Kazakhstan. And the forthcoming Earth Summit in Brazil will see China "negotiate and cooperate with other countries in solving environmental and development problems", according to a statement from the science commission.

The focus on science and technology extends beyond diplomacy to bricks and mortar. For example, construction is well under way at the Shangdi Information Industry Base in the western suburbs of Beijing. Part of the city's high-technology development zone, the complex covers 1.8 square kilometres and, when completed, will have room for more than 100 companies.

Peter Gwynne

Profit motive faces stiff challenge

Washington

RUSSIAN weapons scientists are being asked to do something that their Western counterparts have done poorly, if at all — use their knowledge to stimulate their country's civilian economy.

The international science centre that Western nations are setting up in Moscow (see *Nature* 355, 756; 1992) is intended primarily to keep Russian weapons makers from going to work for potentially hostile countries by providing them with meaningful collaborations in science and technology (see accompanying story). But Robert Gallucci, a US Department of State official, told the Senate Foreign Relations Committee last week that another important goal is "to contribute to the process of converting the former Soviet command economy to a market economy".

Such a conversion would require the 20,000 or so scientists and engineers working at Russia's two major nuclear weapons laboratories to pursue fundamentally new areas of research. While Western officials say they hope to enlist the Russians in programmes involving nuclear reactor safety, environmental cleanup and technology that could lead to commercially viable products, these same officials admit that the Russians have no

history of carrying out such work.

"It's hard to imagine how we will pull it off," says Ed Dowdy, a nuclear engineer with 20 years' experience in the weapons field now serving as science adviser to Gallucci, "but we're certainly going to try."

The United States and the European Community have each pledged \$25 million towards the centre, which was pro-

expected to reach nearly \$100 million.

The governments involved in the centre hope that, through collaborations with the West, the Russian weapons makers will turn their country's swords into ploughshares. But they face formidable obstacles.

The first is that turning weapons technologies into products for the consumer is a tricky business. Despite several years of

federal laws, directives from the Bush administration and the active support of the US Department of Energy to do exactly that, the major nuclear weapons research facilities in the United States — the Lawrence Livermore, Los Alamos and Sandia national laboratories — have so far had little success.

For example, the tool preferred by the US government to forge such ties with industry, called a Cooperative Research and Development Agreement (CRADA), has been barely used. Livermore announced its first CRADA only six weeks ago and Los Alamos has signed four such agreements, all within the past year. Sandia, which is de-

voted to the engineering aspects of making weapons and conducts little pure research, has a slightly better track record, with some 18 CRADAs between itself and various companies.



Visiting Arzamas, Sig Hecker (front row, 2nd from left) and John Nuckolls (front row, far right) pose with their weapons laboratory colleagues.

posed only last month and is expected to be open by the early summer. Japan and Canada have also promised to contribute to a fund that, bolstered by money from industry and non-profit organizations, is

Politicians to call funding shots at the centre

PEER review may be the hallmark of selecting worthwhile scientific projects in the United States and elsewhere. But do not look too hard for it at the new international science centre being set up in Moscow to employ Russian weapons makers and keep them from slipping off to work for potentially hostile countries.

Politicians will decide who works with the Russians after panels of scientists have eliminated ideas that lack sufficient technical merit. Moreover, many of those ideas will come from the governments that are putting up the money to fund the centre.

In testimony last week before the US Senate Foreign Relations Committee, US officials made it clear that their chief worry about the millions of dollars flowing into Russia's nuclear weapons laboratories is whether the money is going to the right people and whether they can watch it being spent. "As a member of the governing board of the centre, the US will be in a position to ensure that a substantial portion of the projects sponsored by the centre do, in fact, directly engage the weapons scientists and that all centre projects involve the necessary degree of transparency", explained Robert Gallucci of the US State Department. "The latter point is important to ensure adequate financial and programmatic monitoring of centre projects."

A recent report by the National Academy of Sciences (*Nature* 356, 181; 1992) urges the US government to create a director's fund so that staff can independently spend a portion of the centre's budget on what it thinks are the best ideas. Without such authority, the report says, the centre "will be seen as a needless and powerless middleman between proposers and funders".

That idea has not caught on within the Bush administration. Qualified scientists from around the world will be asked, by fax machine, to comment on proposals that have been submitted, according to Ed Dowdy, science consultant to Gallucci. But the awards, he says, will be made by "people of international stature, not broadly based in science, who are capable of making good decisions". Giving the director of the centre the authority to make awards "won't happen", he adds.

Gallucci told the US Senate that he has received more than 100 proposals to employ the weapons scientists since the centre was proposed scarcely a month ago. Most have come from government agencies, such as the Department of Energy. The proposals need not specify which scientists they wish to employ, Dowdy says; those that do not will be matched up with an appropriate research team.

J.D.M.

Another problem is the total absence, until recently, of any private sector in the republics of the former Soviet Union with which to interact. Because the scientists have been deprived of such commercial interactions, it is impossible to know with any accuracy the types of projects for which they might be well suited.

Sig Hecker, director of Los Alamos, and John Nuckolls, director of Livermore, visited their Russian counterparts late last month in Arzamas-16 and Chelybinsk-70, which previously had been closed to Western visitors. Two weeks earlier the Americans had played host to the directors of the two Russian laboratories.

Speaking to reporters last week after appearing before a closed committee of the US House of Representatives, Hecker called the Russian weapons laboratories "world-class facilities" with excellent research programmes in such areas as explosively driven high magnetic fields and inertial confinement fusion. But he pointed out that they have no experience in working with the civilian sector and no user facilities of the type common in US laboratories, where many companies conduct proprietary research on expensive machines built by the US government.

At the same time, starting from scratch may have some advantages over operating within a bureaucratic system such as that in the United States. "The disadvantage is that they have no private industry," says Nuckolls about his Russian colleagues. "The advantage is that there are no laws that inhibit such interactions."

The need to keep up the arms race during the Cold War has also shaped the research agenda for scientists in the former Soviet Union. Speaking about the need for research to help clean up a 40-year legacy of environmental problems, Peter Hoffman, a science counsellor in the German Embassy in Washington, says that the weapons scientists "are capable of doing so, but in the past there were political reasons not to do it."

The new centre in Moscow is expected to stimulate this economic conversion by matching Western companies and agencies looking for expertise with Russian research teams looking for encouragement and financing. But business prefers to operate within a stable environment, says D. Allan Bromley, science adviser to President George Bush.

"We must resolve such important issues as ownership and intellectual property rights," Bromley told the US Senate committee at its hearing last week. "Without a structure, our companies will be reluctant to play a role in this exercise."

For all those reasons, then, the attempt by the West to involve Russian weapons scientists in a laundry list of fruitful collaborations may fall short of the mark. But their failure will not be for lack of trying.

Jeffrey Mervis

Strasbourg project grows

London

THE United States and Britain have agreed for the first time to contribute to the Human Frontier Science Program (HFSP), the largely Japan-financed research organization based in Strasbourg. At the same time, the two-year-old programme has begun a search for a successor to Sir James Gowans, a British immunologist who will be stepping down as secretary-general in March 1993.

HFSP owes its existence to a Japanese proposal at the G-7 meeting of governments of industrialized states at Venice in 1987 for an international attack on the understanding of key issues in biology, and for the exploitation of that knowledge in other fields. First reactions elsewhere were sceptical, chiefly on the grounds that the Japanese proposals were too vague.

But HFSP has now announced the outcome of its third round of research awards. Trustees (comprising two members from each of the G-7 countries, the European Commission and Switzerland) last week approved 37 research grants worth \$24 million over three years, and a further 128 two-year postdoctoral fellowships, each worth \$42,000 a year.

The budget of HFSP will grow to nearly \$38 million next year from its current \$31 million, helped by \$3.7 million from the United States and a British contribution of £350,000 (US\$600,000). France contributes \$1.6 million a year, partly as a quid pro quo for the Strasbourg location, apparently influenced by the personal intervention of President François Mitterrand, and Switzerland, Italy and Germany also participate. But Japan remains the major contributor, with close on 80 per cent of the total budget.

It was also agreed last week that Gowans, past secretary of the Medical Research Council, will retire next year after serving four years as secretary-general of the programme. He said he agreed recently to accept a one-year extension of his post, but not to seek reappointment. "I want to get back to my garden", he said.

HFSP's immediate problem is that research applications far outnumber awards. In this year's round, roughly 87 per cent of applications were refused. Gowans says that the only solution is a larger budget.

The ground-rules for the annual competitions specify that research proposals should involve interdisciplinary collaboration between researchers from different member countries. They also must fall in one or other of the two research rubrics, brain function and molecular approaches to problems of biological function such as morphogenesis and energy conversion. Postdoctoral awards are made for work in other member countries.

As in previous years, US researchers

seem to have been involved in the greatest number of applications and to have done best in the competition. Nearly a third (12) of the 37 successful applications have group leaders from the United States, while 48 US researchers were involved in the successful applications.

This year, principal applicants from Germany and Japan each secured four awards (out of 29 and 26 applications respectively). France was more successful, with six out of 25 applications, and Britain less so (with two successes in 34



Gowans to stay one more year

applications in which a researcher from Britain was the principal applicant).

Among this year's successful applications is a project to understand the cerebellum. The collaboration will be based at the Brain Research Institute of the University of Zurich and will involve the physics department at the Federal Polytechnic (ETH) and the Electrotechnical Laboratory at Tsukuba. Other projects include a study of the molecular and cellular biology of learning in *Drosophila* and of *Rhizobium* nodulation signals in plant development.

HFSP claims that it stands out among international research organizations in its use of genuinely international peer-review panels, themselves appointed by an international scientific council representing member states. The mechanism resembles that used by the now-defunct SCIENCE programme of the European Commission, but is broader in geographical coverage.

One member of the scientific council believes the system of peer review is "unique". J. Edward Rall of the US National Institutes of Health, now the chairman of the council, says that he was originally sceptical of the project, but that he has been "most pleasantly surprised" by the interest and the quality of research grant applications.

John Maddox

Firing up for Earth summit

São Paulo

BRAZIL'S environmental secretary, José Lutzenberger, was fired last weekend as the government moved to end an internal dispute that officials fear could damage the forthcoming Earth Summit in Rio de Janeiro. Lutzenberger has been replaced by José Goldemberg, a physicist who is already Minister of Education and, now, certainly the most politically powerful scientist in the recent history of Brazil.

Lutzenberger was dismissed by President Fernando Collor de Mello after making derogatory comments about the government's ability to carry out environmental reforms. Last month, at meetings in New York to prepare for the June summit, Lutzenberger praised a decision by the G-7 states to reduce by one-sixth their scientific and environmental aid to Amazonia because, Lutzenberger said, he feared that a corrupt Brazilian government would squander the original US\$1,500 million that had been pledged. And he questioned the integrity of one of the agencies within his ministry, the Institute of the Environment and Natural Renewable Resources (IBAMA). Collor also fired the president of the institute, biologist Eduardo Martins.

The institute has been in turmoil for the past year. Martins had been appointed by Lutzenberger last autumn after the previous head, Tania Munhoz, became embroiled in a battle with Lutzenberger over his style of leadership. Martins offered to resign after being criticized by Lutzenberger, but Lutzenberger refused to accept the resignation. Martins, a primate specialist, says that he is "returning to the monkeys".

Lutzenberger is an agronomist who is allied with the radical environmental move-

ment in Brazil. Since his appointment two years ago he has been in a running battle with those who advocate rapid development of the country's natural resources. His statement last month that the environmental institute was in league with the timber industry reflects the tension between his views and those of Collor's government, which is promoting a policy of "sustainable development".

Goldemberg, who has strong ties to the international scientific community, is also viewed as a savvy politician. He was appointed secretary of science and technology shortly after Collor took office two years ago, and left his deputy in charge when he moved to the education ministry. For a short time this year he also served as Minister of Health after Collor sacked the minister for alleged financial improprieties. Goldemberg will remain as head of both ministries at least through the Earth Summit in June.

Lutzenberger, on the other hand, has shown little aptitude either for management or compromise. Even his friends admit that he has been something of a loose cannon at political gatherings, and they are unhappy that he has ignored remote sensing data showing the increasing deforestation of the Amazon in making the case for environmental preservation to Collor and his fellow Cabinet officials.

Goldemberg has promised to collect the money already pledged to environmental projects in Brazil. Brazil is paying fines on US\$167 million lent by the World Bank, and none of the \$250 million approved by the G-7 group has reached the country. In the past, Goldemberg has demonstrated good relations with such international lending bodies.

Ricardo Bonalume

UK UNIVERSITIES

Better luck second time?

London

HAVING weathered a storm of criticism in its first attempt to rank UK universities by the quality of their research, the University Funding Council (UFC) is taking a different approach the second time around.

Rather than using a anonymous peer-review team to evaluate all department members at each university, as it did five years ago, the UFC will examine only the work of active researchers. Only a researcher's two most significant recent papers will be reviewed and ranked according to the quality of the journal in which it is published. The UFC also plans to name the members of the review committees, who will be appointed next month and will start their evaluations in June. More than 150 institutions of higher education

will be reviewed, and the resulting rankings will form the basis for UK research funding in 1993 and 1994.

The Committee of Vice-Chancellors and Principals and the Committee of Directors of Polytechnics say they are satisfied that the new system of ranking is fair. But the Association of University Teachers is concerned that the general policy of rewarding the past will make it harder for up-and-coming institutes to get their foot on the funding ladder. They say that the system also encourages a separation of research-based departments and teaching-based departments, which runs counter to the traditional view that teaching is carried out most effectively in an atmosphere of research.

Allison Abbott

Budgets cut again

Cape Town

WHILE South Africa rejoiced in last Wednesday's referendum vote in support of an end to white minority rule, the budget, announced the same day, offered little cheer to the research community. Although government expenditure rose by 16 per cent (roughly in line with inflation), the science budget, divided between the five statutory councils, fell by 3.6 per cent.

The annual reductions are the second in succession for South African researchers, who last year saw cuts in the budgets of four of the five councils for the first time in recent memory despite an overall increase in government spending. And the figures are especially disappointing to those whose hopes were raised last autumn by comments from the Minister of National Education, Louis Pienaar, who warned against the perils of cutting science funding in a speech to the Science Advisory Council. Pienaar heads the department through which research funds are channelled.

Most of the large increase (24 per cent) in education funding was awarded to the Department of Education and Training, which is responsible for black schools. Even so, universities fared slightly better than the research councils, receiving increases in their subsidies of 1.6 per cent.

All of the five research councils funded through the science vote received cuts in their parliamentary grants. The 1992-93 allocations to the councils were as follows: the Council for Scientific and Industrial Research (CSIR), US\$69.7 million, down 5.4 per cent; the Foundation for Research Development, \$34.9 million, down 3.5 per cent; the Human Sciences Research Council, \$23.7 million, down 0.5 per cent; the Council for Mineral Technology (Mintek), \$18 million, down 1.3 per cent; and the Medical Research Council, \$13.4 million, down 2.3 per cent.

In addition, the new Agricultural Research Council (ARC), which will become the country's sixth statutory council on 1 April, received a parliamentary grant of \$78.6 million. The new council will comprise ten research institutes, two research centres and one directorate currently falling under the Department of Agricultural Development.

The department's incumbent president, Alec Heyns, says that the creation of a statutory council will give its administrators greater freedom to hire scientists and assign them to specific projects. Unlike departments, councils may also solicit funds from the private sector.

But some scientists are sceptical about the value of the new council. "The play has been changed, but many of the actors are the same", says Roy Siegfried, director of the Percy Fitzpatrick Institute of African Ornithology at the University of Cape Town.

Michael Cherry

Private doubts, public support

Washington

DAVID Kessler, commissioner of the Food and Drug Administration (FDA), is pressing ahead with reforms to the agency's drug approval process despite serious misgivings about some of them.

Documents uncovered by a panel of the US Congress that has oversight of FDA suggest that Kessler and other senior FDA officials disagree with some of the 11 changes proposed last autumn by the White House Council on Competitive-ness, a body chaired by Vice President Dan Quayle. The council's job is to prevent federal regulations from imposing unnecessary burdens on industry (see *Nature* 354, 173; 1991).

Two of the council's recommendations involve a plan to farm out the review of new drug applications (NDAs) to outside groups and to allow local institutional review boards (IRBs), rather than the FDA, to decide when to begin human clinical trials of a new drug. Pharmaceutical companies, which have long complained about the backlog of drugs awaiting approval at FDA, welcome the reforms as a way to reduce approval times for new drugs. But Congressional critics argue that FDA is being held hostage to the political desires of an administration bent on deregulation. And consumer health group advocates fear that the changes will weaken FDA's ability to scrutinize drug applications and compromise patient safety.

Last week, Kessler told a US congressional subcommittee that any notion that he was forced to endorse these new procedures was "nonsense. These are things that I believe in, things that I believe will work", he told the House subcommittee on human resources and intergovernmental relations of the Committee on Government Operations. But preliminary investigations by the subcommittee chairman, Representative Ted Weiss (Democrat, New York), suggest that Kessler's public words are at odds with his private views.

Documents obtained under subpoena from the FDA demonstrate that Kessler was overruled by the White House after complaining about the council's recommendations to hire outside contractors to review new drug applications and to remove FDA from the direct review of some human clinical trials. An undated executive summary of the council's working group on the drug approval process noted that Kessler "strenuously objected" to a proposal that would allow companies to opt for an external review of their clinical data. And, in a letter dated 1 August 1991 to the deputy secretary of the Department of Health and Human Services (HHS), Kessler outlined similar concerns about the external review of NDAs.

"The idea may seem attractive", Kessler

wrote to HHS. "But it has major problems and could well slow down reviews of new drugs, and decrease the quality and credibility of those reviews." Such an outside review would be cumbersome and inefficient, make it difficult for FDA to audit the reliability of drug application data and force FDA to rely too heavily on summaries of data, he explained. He said there are not enough qualified senior scientists to do the reviews, and that those who are available could face a conflict of interest because of their ties to drug companies.

Despite those positions, which Kessler did not repudiate at the hearing, he told the subcommittee that the FDA has begun a pilot study of the idea of contracting out clinical reviews. However, he stressed that the FDA would retain the final decision as to whether a drug is safe and effective and should be approved for marketing. Kessler said that the pilot study would compare the costs of external and in-house reviews and gauge the number of qualified and eligible reviewers. Although the Competitive Council recommended that as many as a dozen drugs receive such an external review over a period of 16 months, FDA says that it will be able to afford only two such reviews this year.

The recommendation that would allow companies to submit applications for phase 1 (initial) human clinical trials to IRBs is also controversial. If implemented, IRBs (usually at a university or hospital), not FDA, would determine whether animal data are sufficient to warrant testing an experimental drug in humans. Such a move, the Competitive Council asserts, would allow FDA to concentrate its attention and resources on drugs that survive the early stages of investigation.

Critics argue that FDA's absence could have the reverse effect and actually slow the approval process. And, while some IRBs are capable of making these judgments, most lack the necessary expertise in medicinal chemistry, pharmacology and toxicology to determine when drug testing in humans could proceed safely. Similar proposals have met with an unenthusiastic response from IRBs.

While Kessler is publicly backing the move, he expressed doubts in a letter to Louis Sullivan, secretary of HHS, dated 22 January. Anticipating that the medical community will support the idea only if IRB approval is limited to small human studies or studies for new indications of approved drugs, Kessler writes, "we [FDA] are concerned that including large studies in humans within the scope of this proposal would subject us to criticism that we are exposing patients to unreasonable risks . . . the IRB community will not want to take responsibility for such trials".

Diane Gershon

Taking Russia's pulse

SCIENCE ministers from the more than 20 countries that belong to the Organization for Economic Cooperation and Development (OECD) have agreed to turn their considerable analytic tools upon their former enemy, the Russian Republic. Meeting earlier this month in Paris, the ministers approved a \$750,000 study of the state of science in that republic, following the model of similar studies of member countries in which the OECD examined the people, facilities, equipment and other components of the research enterprise. A US official estimated that the study would take a minimum of 18 months, and that the United States would pay its usual share of one-fourth of the total cost.

The ministers also approved a proposal from Austria to host a conference on the brain drain that threatens many of the ex-Communist countries of Eastern Europe and the former Soviet Union. The goal of both the conference and the study is to foster the exchange of scientific information and increase cooperation between the West and the new democracies in Eastern and central Europe.

J.D.M.

New icebreaker

US ANTARCTIC researchers are about to get their second major research vessel, the 308-foot Nathaniel B. Palmer, which was launched last week on its maiden voyage to the southern oceans. The new ship's first mission will be to transport US researchers to a station on a floating iceberg in the



The Nathaniel B. Palmer brings NSF's Navy to full strength.

Weddell Sea, where they are studying ocean currents and the global climate. It can carry 37 researchers for up to three months (breaking three-foot-thick ice when necessary) and has more than 6,000 square feet of laboratory space, including advanced acoustical instruments and computer facilities. The US National Science Foundation operates the \$84-million vessel on a long-term lease. With the Palmer joining the veteran 218-foot Polar Duke, NSF finally has two full-time research vessels to complement its three Antarctic bases.

C.A.

British academic morale and pay

SIR — In his letter on "Action on academic pay" (*Nature* 356, 10; 1992) Howard Morris demands that the Committee of Vice-Chancellors and Principals (CVCP) "call for a pay review body for university academic and related staff".

The CVCP has long been in favour of a pay review body and has been campaigning for its establishment since early 1990. The debate in the House of Commons as recently as 3 March showed that the government's long-standing determination to resist this idea continues.

DAVID HARRISON
(Chairman)

*Committee of Vice-Chancellors
and Principals,
29 Tavistock Square,
London WC1H 9EZ, UK*

SIR — In its current evaluation of university and polytechnic departments, the Universities Funding Council (UFC) commits a serious methodological error in attempting to judge research performance by, among other means, the amount of grant income a department is able to attract. The UFC argues that, because the award of a grant is based on independent peer review, the size of a department's grant income reflects the valuation placed on its research plans by the wider academic community.

This choice of yardstick conflicts directly with the UFC's aim, which is "to ensure that resources for research are used to the best advantage". To assess cost-effectiveness, a performance indicator is required that measures number of units of output (for example publications) per unit of input (for example income). Clearly, grant income is not a performance indicator. It is an input measure. As a consequence, the use of grant income to assess performance inevitably produces gross anomalies of measurement.

Suppose that two departments, A and B, produce publications of identical quality and quantity over a given period. Suppose also that A receives twice as much research-grant income as B, this being the only difference in their funding. If grant income is used as a performance indicator, then the department with more grant income will be rated more highly. However, in terms of the goal of identifying cost-effectiveness, it is obvious that the reverse is true. The department with less grant income should be rated more highly, having produced the same volume of research with fewer resources.

The inappropriateness of using an input measure as a performance indicator is further highlighted by the fact that a

department with an outstanding record of research in areas that require little or no external funding would receive a zero rating, whereas a department with many grants would gain a high rating even if it produced no output at all.

Unlike journal peer review, which critically assesses the merit of a finished product, grant-agency peer review can only make a guess about the promise of future work. To claim that the expert opinion of grant-agency assessors in itself constitutes a judgement of value for money is to beg the question entirely. Whether value for money has indeed been obtained can be discovered only by looking at a department's real outputs (not conjectured ones) in relation to its inputs. As per capita grant income is not a performance indicator, it is structurally incapable of answering this question.

A measure based on grant income necessarily results in a substantial amount of double counting that spuriously inflates the performance of grant-rich departments. The same piece of work can receive two-fold credit, first as anticipated output in a research proposal (measured by grant income) and second as actual output at a later date (measured by number and quality of publications).

The adoption by the UFC of a yardstick based on grant income is clearly antithetical to its objective of obtaining value for money. The continued use of income to measure performance leads inexorably to a situation in which inefficient departments are rewarded and cost-effective departments are penalized.

RAPHAEL GILLET

*Department of Psychology,
University of Leicester,
Leicester LE1 7RH, UK*

SIR — Howard Morris draws attention to successive failures by the Committee of Vice-Chancellors and Principals (CVCP) to protect the pay and morale of academic staff in British universities. It is clear that the CVCP has been tested and found wanting in this respect and must be considered no longer fit to negotiate on academic pay.

But let us not forget the role of the government in this sorry saga. In December 1990, through my MP, I raised specifically the issue of recruitment of scientists into research careers with the Secretary of State for Education and Science, Mr Kenneth Clarke. In response to my complaints about the difficulty of attracting young people into science at the postdoctoral, postgraduate or even undergraduate level, Clarke's reply was that he could not agree with my premises and he quoted figures sug-

gesting that *more* people rather than fewer were choosing science as a career. Clearly this is not the present experience of those involved in scientific research and training in the United Kingdom.

With a General Election now imminent it is worth recalling that, unlike vice-chancellors and principals, Ministers of Education and Science are voted into office and can occasionally be held to account for at least some of their actions.

M. J. CLEMENS

*Department of Cellular
& Molecular Sciences,
St George's Hospital Medical School,
London SW17 0RE, UK*

SIR — As a former parliamentary candidate and a former nuclear structure researcher, I should like to point out two issues of science administration that must be tackled during the next parliament.

One is the division between 'big' science and 'little' science. It is not sensible that the same institution should be responsible both for funding such large expenditures as the UK subscription to CERN and other international collaborations and for funding research studentships and the many small experiments carried out in university departments. The budget of the Science and Engineering Council (SERC) is unnaturally inflated by 'big' science, which makes considered decisions by politicians and by SERC itself almost impossible to make. Last year's decision by SERC to impose a 10 per cent across-the-board cut on all panels smells of dereliction of duty. The CERN subscription in particular should be separately accounted.

The other issue is the coming block obsolescence among UK university staff. Because of the rapid expansion of university education in the 1960s, a high proportion of academics now in tenured posts were appointed within a few years of each other and will retire within a few years of each other in the early part of the next millennium. Scientists of first-class abilities who had the misfortune to be born in the late 1940s or early 1950s have found it very difficult to continue in fundamental research. Science's loss might have been industry's gain, but there is little evidence that British industry has taken advantage of this windfall. During the next parliament, arrangements should be made, possibly by the use of temporary teaching posts of between 5 and 15 years duration, to spread the period over which the next generation of university teaching (and perhaps a few full-time researchers?) is recruited.

DAVE PERKIN

*2 Field Close, Bassett Green,
Southampton SO2 3DY, UK*

Taxonomy of taxonomists

Kevin J. Gaston and Robert M. May

A crude estimate of the number of taxonomists shows a striking mismatch between the geographical location of practitioners and biological diversity. At a time when resources are limited, what needs to be done?

THE systematic classification of the diversity of life on Earth may be thought of as beginning in 1758, when Linnaeus recognized some 9,000 species of plants and animals. Today the total number of named species stands at around 1.5 to 1.8 million: the uncertainty about so basic a fact is remarkable, and derives from the absence of centralized and coordinated records for most groups of organisms. How many species there may actually be is much more uncertain, with estimates ranging from 3 million to 80 million or more¹⁻³.

With growing concern about environmental issues, in particular conservation of biological diversity, there is much debate about the level of support for taxonomic research^{4,5}. Much of this debate is hampered, not only by our ignorance about how many species there are, but also by uncertainties about the taxonomy of the taxonomic workforce itself. How are taxonomists distributed among countries? Among taxonomic groups? And is the population of taxonomists in stasis, growing or decaying?

Analytical studies of these questions are few and fragmentary, mainly because the underlying facts have not been put together. Here we venture some rough answers, as well as some opinions about the implications. As will become apparent, some of our estimates are very crude indeed. We nevertheless believe they are a useful first step, if only to goad others toward correcting them.

Taxonomists among taxa

A study by the US National Science Foundation (NSF)⁶ in 1985 estimated roughly 8,000–10,000 taxonomists in North America (this estimate was not cheaply made, but ultimately it rests on responses to a mailed questionnaire, with the roughly 2,500 responses crudely scaled up by a factor of 3–4 or so). We would give more credence to this study's comparisons among taxonomic groups than to its absolute numbers: roughly 30% of US taxonomists are botanists, 5% work on fossils, 2% on microorganisms and 60% or so on animal phyla. Of animal taxonomists, 32% work on tetrapods (vertebrates other than fish), 11% on fish, 32% on insects and spiders ('entomologists'), and 25% on other invertebrates. These proportions are remarkably similar to those in Australia⁷, where a careful study revealed 588

researchers distributed among tetrapods, fish, insects-and-spiders, and other invertebrates in the proportions 32, 6, 30 and 32% (see table).

Assuming these North American and Australian proportions to be representative of the global distribution of taxonomists among animal groups, we can compare them with the numbers of recorded species (measured in thousands) in the four categories: 22 tetrapods, 19 fish, around 1,000 insects-and-spiders, and around 500 other invertebrates.

Taxonomic group	No. of taxonomists (% of total)	Estimated no. of species per taxonomist	Estimated no. of unrecorded species per taxonomist
Tetrapods	190 (32)	17	3
Fish	33 (6)	190	34
Insects and spiders	178 (30)	840	400
Other invertebrates*	187 (32)	700	430
Total	588 (100)		

* Includes crustaceans, molluscs, echinoderms, coelenterates, sponges, and helminths (but not protozoans). Data from B. J. Richardson (personal communication).

Thus, if there are n taxonomists per species for recorded tetrapods, then there are roughly $0.3n$ for fish and around 0.02 – $0.04n$ for the average invertebrate species. Pleasingly, these estimates are concordant with a different study based on average numbers of papers per recorded species, which gives 0.9 for tetrapods, 0.4 for fish, 0.02 for insects-and-spiders, and 0.04 for other invertebrates (see Table 3 of ref. 8). More generally, there are roughly twice as many taxonomists for each of the roughly 270,000 recorded plant species than for each recorded animal species. But allowing for the differences in recorded species richness among animal groups, we see that, with respect to taxonomic attention, the average plant species does about an order of magnitude worse than the average vertebrate species and about an order of magnitude better than the average invertebrate.

The Australian study summarized in the table estimates how many yet unrecorded Australian species there may be in each of the above four categories. The resulting guess is that there are 3 unrecorded tetrapod species per taxonomist,

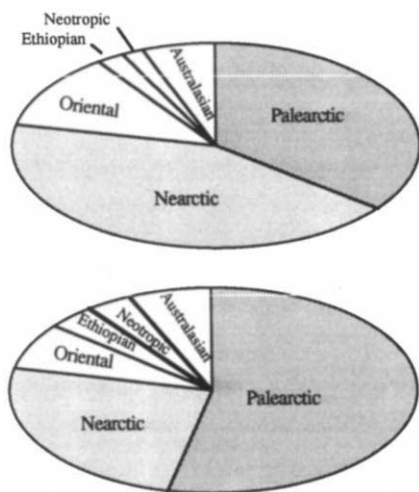
34 unrecorded fish species, and around 400 unrecorded species of invertebrates. That is, there is a full order-of-magnitude discrepancy in the level of effort given to yet-unrecorded fish species versus other vertebrates, widening to two full orders of magnitude for invertebrates. The invertebrate fauna of Australia is likely to be more completely recorded than those of most tropical regions, so the discrepancies among groups on the right of the table are likely to be even wider for global faunas.

Overall, the distribution of taxonomists is ill-matched to the species richness of taxa, and to the magnitude of the jobs remaining to be done for different groups. Insofar as recent initiatives may result in more money for taxonomic research, we believe that because resources are limited and priorities must be set, plants should benefit more than vertebrates, and invertebrates more than either. Half the remaining tropical rainforests will be gone within the next 30–50 years, and as they vanish no one knows how many species of invertebrates, fungi and soil microorganisms will be extinguished, nor what the larger consequences may be.

Distribution of taxonomists

It is difficult to estimate precisely the total number of taxonomists working on any one taxon in any country. Comparisons among countries or among taxonomic groups are perhaps a bit easier because we can hope that the many biases and imperfections in the 'data' are broadly similar from one country to the next or from one group to another. For example, the figure (top) estimates the numbers of practising ecologists⁹, suggesting that around 80% of ecological researchers are based in North America or Europe, with only 4% based in Latin America or sub-Saharan Africa, where there is so much biodiversity.

We have made a crude estimate of how active insect taxonomists are distributed among biogeographical realms by looking at the countries of origin of the roughly 1,200 people from 57 countries who borrowed from the entomological collections at the Natural History Museum in London over the past 5 years. The biases and uncertainties in such an estimate are numerous, but it again indicates that about 80% of insect taxonomists are based in North America



Top, Distribution among Wallace's biogeographical realms of about 16,000 authors of papers listed under 'ecology' in *Biological Abstracts* 1982-83 (after ref. 9). Bottom, Biogeographical distribution of about 1,200 taxonomists borrowing from insect collections at the Natural History Museum between January 1986 and June 1991. Palearctic: Europe and Siberia; Nearctic: North America; Oriental: Near East to Far East and Malaysia; Ethiopian: sub-Saharan Africa; Australasian: New Zealand, Papua New Guinea and the Pacific Islands.

and Europe and around 7% in the neotropical and Ethiopian realms. Similar results emerge from a more limited analysis of about 271 institutions borrowing from the Royal Botanical Gardens, Kew in 1989: of the total of 7,391 species or 'types' which were borrowed, 78% went to palearctic or nearctic institutions, compared to 6% to neotropical or Ethiopian ones (G.L.I. Lucas, personal communication). The mismatch between geographical location of researchers and of biological diversity is again evident.

Of course, many taxonomists based at North American or European institutions deal mainly with tropical groups. Even so, the numbers of people and methods of working are clearly inadequate for cataloguing, much less understanding, diversity in the tropics. For example, at current rates it would take about 380 years to complete the Flora Neotropica, and 950 years to complete the inventory of fungi for the same region¹⁰. We must develop new ways of going about time-honoured tasks.

In some respects these new approaches will seem cruder than traditional methods, for example in the use of 'parataxonomists', who may lack scholarly credentials. In other respects, new approaches need to embrace modern technology in ways as yet barely dreamed of. With appropriate financial support and attitudes, not only could information about previously and newly recorded species be collected and stored on CD-ROM, but efficient programs could be developed for 'keying-out' spe-

cies against a coherently organized database in ways that could revolutionize this time-consuming task. Such databases would have the extra virtue of addressing one of the tensions between researchers in industrialized and developing countries. Using the *Index Herbariorum*, Parnell¹¹ has observed that the workforce of tropical plant taxonomists in universities roughly doubled from 1960 to 1980, almost wholly as a result of new posts in tropical universities. Reference collections cannot be built so quickly, and so North American and European institutions remain the primary centres for such collections. Good use of information technology can make distant collections available readily to tropical researchers.

Somewhat better — but still imprecise — geographical comparisons of levels of effort in entomological taxonomy can be made among North America, the United Kingdom and Australia. The Entomological Society of America (ESA) recently completed a lengthy study, based on an order-by-order survey, of numbers of entomologists in North America¹⁵. Although the survey's summary chapter contains some significant misunderstandings, it reveals something like 880 insect-and-spider taxonomists (which contrasts with 501 in this category who responded to the NSF questionnaire, leading to that study's crudely scaled-up estimate of 1,500-2,000 entomological taxonomists in North America). Of these estimated 880, almost 40% came from one order, the economically and medically important Diptera (flies); the other major orders, Coleoptera (beetles), Hymenoptera (ants, bees, wasps) and Lepidoptera (butterflies and moths), account for roughly 18, 11 and 6%, respectively.

Our similar but cruder order-by-order survey of British insect-and-spider taxonomists gives an estimated total of around 200 individuals, of whom 92 are professional taxonomists or systematists partitioned among Diptera, Coleoptera, Hymenoptera and Lepidoptera as 16, 14, 17 and 13%, respectively.

Scaled against the total populations of these three countries, these results give around 4 insect-and-spider taxonomists per million of population in North America, around 4 per million in the United Kingdom, and around 10 per million in Australia. Thus North America and the United Kingdom are similar; it not clear to us whether the Australian number is higher because the underlying count is more complete or because Australia invests more in taxonomy.

Trends in the workforce

So far, we have dealt with a blurred snapshot in time. To deal with trends, we need movies, and for the taxonomy of taxonomists these are even blurrier.

Some studies¹³ suggest that funding for, and overall numbers of, taxonomists and systematists have not so much declined in recent decades, as remained roughly constant while funds and researcher numbers have increased markedly in other areas. But the research component of taxonomy and systematics in the United Kingdom has decreased by 6% in real terms (from £17.5 million in 1980 to £16.5 million in 1990), with a concomitant 7% drop in the total number of taxonomic researchers (from 552 in 1980 to 514 in 1990, an increasing fraction of them on short-term contracts¹⁴).

In the universities, which breed the next generation of taxonomists, the subject is in absolute, not relative, decline. The average proportion of undergraduate time devoted to systematic biology has fallen from 17% in 1980 to 9% in 1990, although the number of postgraduate research students has remained roughly steady over the same period⁴. Probably the most straightforward symptom of stasis or decline in a population or profession is an 'over-aged' age profile: Parnell's analysis¹¹ shows that fewer than one in seven (13%) of tropical plant taxonomists in European universities are less than 40 years old.

The NSF study⁶ of the taxonomic community reported the proportions of respondents in the five categories "established professional, postdoctoral, doctoral student, master's student, other" as 85, 3, 8, 2 and 2, respectively. In 1980 in the United Kingdom, 23% of teachers of systematic biology were 35 and under and 43% were 46 and over; in 1990, the corresponding proportions were 8 and 63%. If we found these demographic trends in a newly discovered species of lemur, we would bring specimens into a zoo and start a programme of captive breeding. But if these trends continue among taxonomists and systematists, how soon will it be before we cannot recognize a new species of lemur? □

Kevin J. Gaston is at The Natural History Museum, Cromwell Road, London SW7 5DB; Robert M. May is in the Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK.

1. May, R. M. *Phil. Trans. R. Soc. B* **330**, 293-304 (1990).
2. Stork, N. E. *Biol. J. Linn. Soc.* **35**, 321-337 (1988).
3. Gaston, K. J. *Conserv. Biol.* **5**, 283-296.
4. House of Lords Select Committee on Science and Technology *Systematic Biology Research* (HMSO, London, 1992).
5. Gee, H. *Nature* **355**, 488 (1992).
6. Edwards, S. R., David, G. M. & Nerling, L. I. *The Systematics Community* (Museum of Natural History, Lawrence, Kansas, 1985).
7. Richardson, B. J. *Search* **14**, 320-323 (1984).
8. May, R. M. *Science* **241**, 1441-1449 (1988).
9. Golley, F. B. in *Trends in Ecological Research in the 1980's* (eds Cooley, J. H. & Golley, F. B.) 1-4 (Plenum, New York, 1984).
10. Prance, G. T. & Campbell, D. *Taxon* **37**, 519 (1990).
11. Parnell, J. *Conserv. Biol.* (in the press).
12. Kosztarab, M. & Schaefer, C. W. (eds) *Systematics of the North American Insects and Arachnids: Status and Needs* (Virginia Polytechnic Ins. Blacksburg, 1990).
13. Slater, J. A. *Bull. ent. Soc. Am.* **Spring**, 4-7 (1984).

The endless search for primality

Is there more to the continuing search for prime numbers than the curiosity that sustains the enthusiasm of stamp collectors? The discovery of the largest so far (at Harwell in Britain and by Cray Research) refers.

THE discovery of the largest prime number so far, herewith made public, may not set the world alight, but is well worth modest celebration, if only because of the circumstances of its discovery. Briefly, it appears that Cray Research provides its customers with a program for telling whether numbers are true primes. Naturally, a purchaser who has just written a cheque for several million dollars will be comforted if he can check that his new purchase verifies the primality of a known prime number, perhaps one chosen at random from a list of known primes.

In this case, the customer is the Harwell Laboratory of what is now called AEA Technology, previously the principal research laboratory of the UK Atomic Energy Authority. The computer is a Cray-2 machine equipped with four central processors. Like others before them, the Harwell people set out to test the primality of large Mersenne numbers, which have the form $2^q - 1$, where q is itself a prime number. What they have found is that the Mersenne number with $q = 756,839$ is also prime. It is much larger than that which previously held the record, with $q = 216,091$.

The sheer size of these numbers is awesome. In decimal notation, Harwell's new prime consists of 227,832 digits, almost exactly three times as many as the previous largest prime. A little mental arithmetic will show that it will never be published as ink on paper: if crammed with text, this page would hold 8,000 characters, so that Harwell's prime would occupy close on 30 pages of *Nature*. And, given that proof-reading would be so nearly impossible as to be pointless, only machine conversion from electronic signals into print would be practicable. (Similar problems arise with the nucleotide sequences of DNA, of course.)

But the discovery should not be scorned on that account. For one thing, the search for prime numbers has an abiding interest, not least because of the continuing uncertainty about their placement among the integers. Euclid, some time ago, showed that the number of primes must be infinite by the simple remark that adding to or subtracting 1 from the product of all consecutive prime numbers up to a certain maximum will yield two further prime numbers, which also goes to show why prime numbers often occur in pairs. (The product of the first three primes — 2, 3 and 5 — is 30, and both 29 and 31 are primes.) In other words, given any sequence of

consecutive primes, it is always possible to construct two that are larger.

But the density of primes diminishes as integers become larger. Handwaving shows that. The larger a number, the larger the number of prime numbers less than or equal to its square root and thus the greater the chance that it will be divisible by one of them. More precisely, according to Poussin in 1899, the number of primes less than a number N is, to a first approximation, $N/\ln N$, where "ln" stands for the natural logarithm. Taken literally, this means that the density of primes declines only slowly with increasing size. For example, the number of primes per thousand consecutive integers in the region of 10^9 would be related to that in the region of 10^6 in the ratio of 6:9, or 2:3.

It is also important that this approximation is only the first term in a slowly converging series, whose accuracy has never been stringently tested. One of the uses of very large prime numbers is that they may, when accompanied by a knowledge of the neighbouring primes, make it possible to test the Poussin formula and other guesses by people such as Riemann at the density of primes. But the likelihood of that will hang on people's energy, and the spare time of machines such as the Cray-2.

Meanwhile, the use of Mersenne numbers in the search for ever-larger primes seems unavoidable, and that is a remarkable story in itself, best told in Volume 2 of Donald E. Knuth's *The Art of Computer Programming*. Marin Mersenne was a monk who first showed in 1644 that numbers of the form $2^q - 1$ are prime only if q is prime (for otherwise the number is divisible by a smaller Mersenne prime) and who conjectured — nobody knows how — that the first prime numbers of this form are those in which q is one of 2, 3, 5, 7, 13, 17, 19, 31, 67, 127 and 257.

Knuth, who is a terrier for these things, records that by the 1880s, it had been shown that $2^{67} - 1$ is not a prime and, soon afterwards, that $2^{61} - 1$ is a prime, which led Mersenne's supporters to plead that he had mistranscribed "61" and "67". But then, unhappily, it turned out (in 1911) that $2^{89} - 1$ is a prime not on Mersenne's list, and (in 1914) that $2^{107} - 1$ is another. Then, in 1922, it turned out that $2^{257} - 1$ is not a prime.

The usefulness of Mersenne numbers in the search for larger primes now rests on a different property of integers of the form $2^q - 1$, that by a suitable interpretation

of the representations of numbers by strings of binary digits ("0" or "1"), division by those numbers can be reduced to shift and complementation operations. Cray Research's diagnostic prime program is based on what is called the Lucas-Lehmer test for the primality of Mersenne numbers, based on a rigorously proved algorithm.

The procedure sounds more complicated than it is. If q is the exponent of the Mersenne number to be tested, the trick is to compute a sequence of numbers L_n by the rule that $L_{n+1} = (L_n^2 - 2) \bmod (2^q - 1)$. The initial condition is that $L_0 = 4$ and the test is embodied in the assertion that the Mersenne number is prime if (and only if) L_{q-2} is zero. The notation "mod" is simply the remainder after division by the Mersenne number. It is quite fun to work out a few examples with pencil and paper; Knuth gives a formal proof.

The hard work, even for a Cray machine, is the repeated calculation of squares — in this case, 756,839 of them. Quite soon in the iteration, the numbers L will be comparable in size with the Mersenne number under test, so that the computer space required to store the square requires twice as many digits, getting on for half a million (in decimal notation).

Making this efficient has been the work of David Slowinski, the computer scientist at Cray Research who designed the program. He says that he has used all the tricks of his trade, and that it nevertheless took nineteen hours of time on Harwell's Cray-2 to obtain a proof that $2^{756,839} - 1$ is prime. But there is a bonus: this Mersenne number (like 31 others) can be used to generate what is called a perfect number by multiplying it by the factor 2^q . A perfect number (which cannot be a prime) is one that is equal to the sum of its prime factors.

It is understandable that the people at Harwell and Cray are delighted with this outcome. It is a triumph of what Knuth would call the programmer's art. But is it more than stamp-collecting? That depends. The obvious need now is for a means of calculating primes in the neighbourhood of giants such as Harwell's, and then for using these to test the conjectures about the density of primes. And while the occurrence of primes may be peculiar, it seems that nobody has yet proved that a systematic procedure for their prediction is impossible. That should keep the armchair practitioners of modular arithmetic happy for years.

John Maddox

Feeding the beast

David Sanders

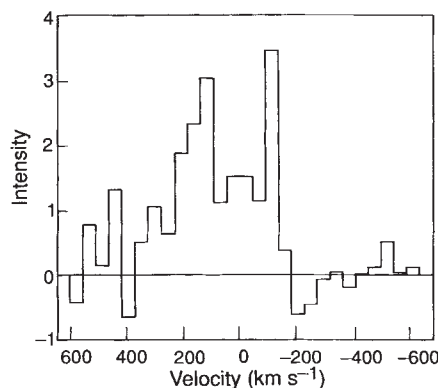
THE exceptionally luminous and distant infrared galaxy IRAS 10214+4724 is again in the headlines, this time following the discovery a few months ago that it harbours an enormous quantity of star-forming molecular gas as indicated by the detection of submillimetre emission from carbon monoxide gas. On page 318 of this issue¹, Solomon, Radford and Downes speculate on the nature of the energy source in 10214+4724, and discuss the implications of the CO detection in the context of nearer ultraluminous infrared galaxies. The authors calculate that the total mass of molecular gas is equal to $2-6 \times 10^{11}$ solar masses, comparable with the total mass of stars in large nearby spiral galaxies. They speculate that the gas mass is a large fraction of the total mass of 10214+4724, as much as 90 per cent, and suggest that the term *primaeval* best fits this object, as it has only just begun the process of converting its initial mass of gas into stars.

The discovery of the high-redshift IRAS galaxy was reported in *Nature* last year by Rowan-Robinson *et al.*² and was the subject of an accompanying News and Views article³ which framed the initial debate on the nature of this exceptionally luminous beast — hidden quasar or protogalaxy? The enormous infrared luminosity, equivalent to that of 3×10^{14} Suns or approximately 30,000 times the total luminosity of the Milky Way, makes this object a rival to the most luminous quasars. Although quasars emit mostly at optical, ultraviolet and X-ray wavelengths, they provide a plausible energy source for the exceptional luminosity of 10214+4724, and one need only invoke a heavy dust shroud to convert the short-wavelength radiation to the longer wavelengths characteristic of the thermal infrared.

The fact that 10214+4724 rivals the luminosity of the most luminous quasars is made more palatable given that we see it at an epoch when the number density of quasars peaks⁴ and the most luminous objects are more likely to be found. However, another possibility for explaining the luminosity of 10214+4724 comes from exploring the implications of the large amount of dust that is needed to produce the observed infrared emission. Rowan-Robinson *et al.* estimate the presence of 4×10^8 – 10^{10} solar masses of dust, depending largely on the shape of the still unmeasured submillimetre spectrum at wavelengths longer than 100 micrometres. Taking the upper end of this range, and assuming a gas-to-dust ratio of 100 — typical of normal spiral

galaxies — the galaxy's total gas mass could easily be in the range 10^{11} – 10^{12} solar masses, equivalent to the total mass (primarily in stars) of nearby large spiral galaxies. If the gas is primarily in the star-forming molecular phase then it would be conceivable that it is feeding a super burst of star formation that is the primary power source.

Of course, all the speculation concerning the nature of the energy source in 10214+4724 could have been for naught



Spectrum of CO (3→2) emission reported by Brown and Vanden Bout⁵. The detection required 17 hours of integration at the NRAO 12-metre telescope (Kitt Peak). The velocity scale corresponds to gas motion in the object rest frame. The observed centre frequency at zero velocity was 105.2331 GHz corresponding to the CO(3→2) rest frequency (345.7959 GHz) redshifted by $(1+z=3.2867)$.

if the source had been incorrectly identified; the original paper² seemed convincing, but the high-redshift ($z=2.286$) source identified with 10214+4724 was actually just outside the IRAS position error ellipse, and one could have argued that the infrared emission had nothing to do with the distant source. However, any such notion quickly evaporated with the announcement by Brown and Vanden Bout⁵ of the detection of an emission line (see figure) from the (3→2) rotational transition of the CO molecule at a redshift of 2.2867 corresponding to the reported optical redshift of 10214+4724.

Although CO emission has become a standard method of determining the mass of molecular gas in galaxies, Brown and Vanden Bout's detection was shocking — the redshift exceeded the largest previously reported CO emission-line redshift by more than a factor of 10, and at first glance the emission was unexpectedly strong, implying the presence of an enormous quantity of molecular gas.

Solomon, Radford and Downes, in this issue, use the CO luminosity to

compute a molecular gas mass of $3-6 \times 10^{11}$ solar masses. They point out that the object's relative infrared and molecular gas properties closely resemble those of nearer ultraluminous infrared galaxies⁶, and note that the choice between starburst and quasar models for its primary luminosity source parallels the debate over luminous infrared galaxies and quasars in the local Universe. But they still leave the question open.

My own view is that in this galaxy and in nearer ultraluminous infrared galaxies, luminous starbursts and quasar activity are triggered together when gas-rich galaxies come close together or even merge. In this picture, the peak in the quasar space density distribution near a redshift of 2 corresponds to the epoch when mergers of extremely rich disk galaxies are frequent. Each merger results in vast quantities of molecular gas being drawn into the merger nucleus, where it fuels a super starburst; the stars and gas then build or fuel the quasar nucleus.

The infrared phase is then plausibly the initial stage of this process, which eventually gives way to a phase of excess optical and ultraviolet emission as the dust shroud is driven away to reveal the nuclear engine. Star formation and disk building presumably began at an even earlier epoch, in which was created the substantial quantity of dust seen in 10214+4724.

Further studies of the object should provide additional fuel for the debate about starbursts and quasars. Draper *et al.*⁷ report that the integrated optical light is strongly polarized (16 ± 2 per cent) for an unresolved object, consistent with the hidden-quasar interpretation; and Soifer *et al.*⁸ report that the equivalent width of the hydrogen emission line ($H\alpha$) is consistent with the largest values found in quasars. But the presence of $3-6 \times 10^{11}$ solar masses of star-forming molecular gas would seem to provide sufficient fuel for those who favour the starburst model. The discovery of additional similar objects would certainly help, but for that we will have to await the more sensitive Infrared Space Observatory mission scheduled for early 1994. □

David Sanders is at the Institute of Astronomy, University of Hawaii, 2680 Woodlawn Drive, Honolulu, Hawaii 96822, USA.

1. Solomon, P. M., Radford, S. J. & Downes, D. *Nature* **356**, 318–319 (1992).
2. Rowan-Robinson, M. *et al.* *Nature* **351**, 719–721 (1991).
3. Lilly, S. J. *Nature* **351**, 697–698 (1991).
4. Schmidt, M. & Green, R. F. *Astrophys. J.* **269**, 352–374 (1983).
5. Brown, R. L. & Vanden Bout, P. A. *Astr. J.* **102**, 1956–1959 (1991).
6. Sanders, D. B. *et al.* *Astrophys. J.* **328**, L35–L39 (1988).
7. Draper, P. *et al.* *Gemini* **34**, 14 (1992).
8. Soifer, B. T. *et al.* *Astrophys. J.* **381**, L55–L58 (1991).

Conviction by genetics

Tony Pawson

MANY otherwise diverse cytoplasmic signalling proteins, molecules which have been implicated as the targets of activated receptor protein-tyrosine kinases, have two domains in common — SH2 and SH3, SH standing for Src homology¹. On page 340 of this issue², Clark, Stern and Horvitz provide persuasive evidence, derived from the fruitful field of nematode genetics, that SH2 and SH3 are indeed essential elements of cellular signal transduction pathways.

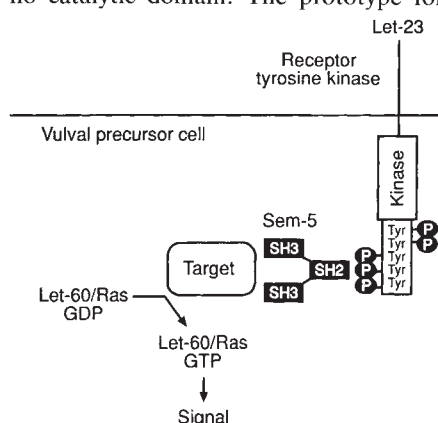
A number of the extracellular signals that control cellular proliferation, migration and differentiation operate through cell-surface receptors with cytoplasmic tyrosine kinase domains. Binding of an external ligand to these receptors induces their dimerization, enzymatic activation and intense autophosphorylation on tyrosine residues, apparently by an intermolecular mechanism. But what are the immediate targets of these activated tyrosine kinases, which in turn might mediate their effects on gene expression, cytoskeletal architecture, cell-cell interactions and intracellular metabolism?

Pointers to the answer came from the observation that the autophosphorylation of mammalian growth-factor receptors induces their association with several cytoplasmic proteins, which bind to specific tyrosine phosphorylated sites within noncatalytic regions of the receptors³. These receptor-binding proteins, which include phospholipase C- γ (PLC- γ), phosphatidylinositol (PI) 3'-kinase, Ras GTPase-activating protein and Src-family tyrosine kinases, are themselves regulators of intracellular signalling pathways, and hence are logical receptor targets⁴. Their association with receptors generally correlates with their phosphorylation on tyrosine and, for PLC- γ , PI 3'-kinase and Src, with their activation. Although they are otherwise entirely different in both structure and function, these cytoplasmic signalling enzymes all share two distinct sequences of about 95 and 45 amino acids, the SH2 and SH3 domains⁵.

Not surprisingly, the conserved SH2 domains are responsible for complex formation with autophosphorylated receptors⁵. The SH2 domains bind weakly to phosphotyrosine⁶, and strongly to phosphotyrosine embedded within a short, specific peptide sequence⁷. They can be aptly described as protein modules, in that they retain their binding properties when expressed individually in bacteria, and normally function in a variety of different polypeptide settings.

The role of SH3 domains has been more enigmatic. They have been inde-

pendently identified in a number of proteins that associate with and modulate the cytoskeleton^{1,8}, suggesting an involvement in cytoskeletal interactions. However, their frequent location adjacent to SH2 domains implies a synergistic activity involving both elements. This notion is bolstered by the identification of several small proteins that contain only SH2 and SH3 domains, and, unlike SH2/SH3 proteins such as PLC- γ , have no catalytic domain. The prototype for



A model for Sem-5 function. The protein is envisaged to be an adaptor, coupling the Let-23 receptor tyrosine kinase to a target that regulates Let-60/Ras guanine nucleotide-binding. In this scheme, the SH2 domain of Sem-5 contacts a tyrosine-phosphorylated site in the tail of the Let-23 receptor tyrosine kinase, while the Sem-5 SH3 domains interact, directly or indirectly, with a Ras regulator, possibly a guanine nucleotide-exchange protein. In outline, the model parallels that proposed for interactions of the heterodimeric PI 3'-kinase with mammalian growth-factor receptors.

this class of SH2/SH3 protein is the product of the *v-crk* oncogene⁹. Another example is the Nck protein, which contains three SH3 domains followed by an SH2 domain¹⁰. One possible function of these proteins is to act as regulatory subunits that couple receptor tyrosine kinases to catalytic targets that themselves lack a receptor-binding SH2 domain. In fact, PI 3'-kinase exploits such an arrangement — it is composed of an 85 K receptor-binding subunit, containing SH2/SH3 domains, associated with a 110 K polypeptide, presumed to be the catalytic subunit¹¹.

From these data, a strong case can be made that SH2/SH3 domains regulate cellular signal transduction. Thus far, however, it has been a matter of guilt by association. Clark *et al.*² have now produced genetic evidence that ensures a conviction. The authors isolated a gene (*sem-5*) from the nematode worm *Caenorhabditis elegans* which encodes a

228-amino-acid protein composed almost exclusively of SH2 and SH3 domains, in the order SH3-SH2-SH3. The *sem-5* gene was identified genetically because it is required for sex myoblast migration, differentiation of cells within the vulval lineage, and larval survival. The recessive vulvaless phenotype displayed by animals with severe disruption of the *sem-5* gene is of particular interest. In vulval induction, a gonadal cell (the so-called anchor cell) transmits a signal to vulval precursors, three of which respond by differentiating along the vulval pathway. Loss-of-function mutations within the genes controlling these signalling events block induction of the precursors, giving a vulvaless phenotype; conversely, gain-of-function mutations can induce inappropriate differentiation of three additional cells, and hence a multivulval phenotype.

Two genes required for signal transduction in vulval cells have previously been characterized in some detail — *let-23* encodes a transmembrane tyrosine kinase very similar to the receptor for epidermal growth factor (EGF)¹², which presumably is the receptor for the anchor-cell signalling molecule; and *let-60*, which acts downstream of *let-23*, encodes a Ras protein¹³. These results are conceptually satisfying, because tyrosine kinases have been shown in mammalian cells to signal through Ras, a small guanine nucleotide-binding protein that is biologically inactive in the GDP-bound state, and is activated by exchange of GDP for GTP. The genetic properties of *sem-5*, and the structure of its product, imply that it communicates between the Let-23 tyrosine kinase and Let-60/Ras. One can readily envisage how Sem-5 might interact with Let-23, by proposing that the SH2 domain of Sem-5 binds a tyrosine phosphorylated site within the activated Let-23 receptor. The carboxy-terminal tail of Let-23 contains several potential autophosphorylation sites, one of which, tyrosine-1242, is in a sequence context (EDNSYLIP) rather similar to that of tyrosine-992 in the EGF-receptor (DADEYLIP), whose phosphorylation induces high-affinity binding of PLC- γ SH2 domains⁷. Consistent with this notion, an allele of *sem-5*, in which an SH2 serine residue required for phosphotyrosine binding⁶ is substituted by asparagine, gives a partial vulvaless phenotype.

The mechanism by which the Sem-5 protein might, in turn, regulate Ras is unknown. One possibility is that Sem-5 acts as an adaptor to couple the Let-23 tyrosine kinase to a target that directly modifies Ras, such as a guanine nucleotide-exchange protein (see figure). If the SH2 domain of Sem-5 binds to Let-23, it seems likely that one or both of the surrounding SH3 domains, which

comprise the rest of the Sem-5 protein, might contact this postulated target. Remarkably, a single amino-acid substitution in the amino-terminal SH3 domain of Sem-5 gives a recessive vulvaless phenotype, and is lethal to young larvae. Clark and colleagues point out that both SH3 domains are required for normal Sem-5 function, because a substitution in either one alone gives a mutant phenotype. These findings bring SH3 domains to centre stage, and provide support for a model in which SH2 and SH3 domains can act in concert to activate signalling pathways.

The majority of nematodes with mutant *sem-5* alleles are defective in several signalling systems, including vulval development and migration of the sex myoblast. Mutations in *sem-5* also suppress the recessive *clr-1* phenotype, which is characterized by sterility and makes cells more visible. This phenomenon might be explained if the *clr-1* defect results in elevated tyrosine phosphorylation, for example by affecting a tyrosine phosphatase. Under these circumstances, inactivation of a critical tyrosine kinase target, such as Sem-5, might suppress the original phenotype. Together, these results imply that Sem-5 is broadly involved in signalling downstream from receptor tyrosine kinases.

What of the future? The plethora of proteins containing SH2 and SH3 domains suggests that there is a complex network of protein-protein interactions that regulate the activation of signalling pathways by tyrosine kinases. Understanding these associations will require structural analysis of SH2 domains bound to phosphorylated peptides, and of SH3 domains; biochemical and molecular identification of SH2- and SH3-binding proteins; and genetic investigation of the biological functions of SH2/SH3. The payoff may be the development of ways to manipulate the signalling pathways that control cell growth and development. □

Tony Pawson is in the Department of Molecular and Medical Genetics, University of Toronto, and the Samuel Lunenfeld Research Institute, Mt Sinai Hospital, 600 University Avenue, Toronto, Ontario M5G 1X5, Canada.

The Earth's core in a nutshell

J. A. Jacobs

UNTIL now, the thermal history of the Earth's inner core has defied mathematical analysis. But on page 329 of this issue¹, Buffett *et al.* present a simple analytical expression that will allow Earth scientists to calculate for the first time this, literally and figuratively, most central of issues in the Earth's evolution.

In 1986 the American Geophysical Union celebrated the fiftieth anniversary of the discovery of the Earth's inner core by the Danish seismologist, Inge Lehmann². That the Earth had a core of higher density than the surrounding mantle had been suggested as early as the nineteenth century, but it was not until 1906 that Oldham³ established its existence. Because seismic shear waves have never been observed to pass through the core, it was assumed to be fluid. The suggestion that there is a solid inner core seems to have been made first by Birch⁴. Since then much work has been done on the constitution of the Earth's deep interior, particularly on conditions at the core-mantle boundary.

The constitution of the Earth's core is critically dependent on the thermal history of the Earth. Unfortunately, temperatures in the Earth's deep interior, 5,000 km below our feet, are among the least well-known of its physical properties. Lord Kelvin⁵ argued at the turn of the century that the heat flow through the Earth's crust was incompatible with an age of the Earth greater than about 25 million years. That this estimate was more than two orders of magnitude too small is because he did not consider solid-state convection in the mantle and because radioactivity had not been discovered.

One of the main reasons for interest in the Earth's core is that the Earth's magnetic field is believed to result from dynamo action in the fluid outer core. The outer core is a good conductor of electricity (it is mainly iron) and is a fluid in which motions can take place — it permits both mechanical motion and the flow of electric current, the interaction of which can generate a self-sustaining magnetic field.

There are two main contenders for the driving force of the geodynamo — thermal convection in the fluid outer core and the growth of the inner core. Freezing of material at the inner core boundary would separate a heavy fraction (mainly iron), leaving behind a lighter fraction in the outer core that would be buoyant and lead to compositionally driven convection (see, for example, Gubbins *et al.*⁶). Light material

rising up will drive fluid motions directly and all of this potential energy must be converted to heat via dissipative processes before the heat escapes to the mantle. Electrical heating is probably the main dissipative process in the core. Heat-driven convection on the other hand is much less efficient — most of the energy is simply carried away by hot fluid rising to the top of the core. Thus the growth of the inner core may play an important part in the maintenance of the Earth's magnetic field.

Rocks as old as 3,500 million years have been found to possess remanent magnetization, so that it is extremely probable that the Earth already had a molten outer core about the same size as that at present. Other studies also indicate that the Earth's core formed either during accretion of the Earth or very soon after (within a few hundred million years). The time of formation and growth of the inner core is a much more difficult question and one to which Buffett *et al.* address themselves.

Their method is based on global heat conservation; that is, the net heat flux from the outer core is equated to that lost from the liquid outer core together with that produced by the growth of the inner core. They assume that temperatures throughout the outer core are determined by the solidification temperature alone (treated as a function of pressure and hence of depth). One advantage of their method is that they obtain an analytical solution which enables the importance of the different physical parameters to be readily assessed.

The authors consider two limiting cases, in which the inner core is thermally perfectly insulating and perfectly conducting. This circumvents the need to know the temperature in the inner core and places bounds on their solution. In the event it appears that there is little difference in their solutions for these two extreme cases. The radius of the inner core is determined by the solidification temperature, which uniquely determines the temperature throughout the outer core. Their solution depends on the magnitude and time dependence of the heat flux across the core-mantle boundary. They illustrate their solution with probable maximum and minimum values of the heat flux and obtain the time for the inner core to grow to its present size as 1,000 million years and 3,600 million years, respectively. An intermediate value of the heat flux across the core-mantle boundary gives a time of 1,600 million years. It would be interesting

1. Koch, C. A., Anderson, D., Moran, M. F., Ellis, C. & Pawson, T. *Science* **252**, 668–674 (1991).
2. Clark, S. G., Stern, M. J. & Horvitz, H. R. *Nature* **356**, 340–344 (1992).
3. Kazlauskas, A. & Cooper, J. A. *Cell* **58**, 1121 (1989).
4. Cantley, L. C. *et al.* *Cell* **64**, 281–302 (1991).
5. Anderson, D. *et al.* *Science* **250**, 979–982 (1990).
6. Mayer, B. J., Jackson, P., Van Etten, R. & Baltimore, D. *Molec. cell. Biol.* **12**, 609–618 (1992).
7. Rotin, D. *et al.* *EMBO J.* **11**, 559–567 (1992).
8. Chenevert, J. *et al.* *Nature* **356**, 77–79 (1992).
9. Mayer, B. J., Hamaguchi, M. & Hanafusa, H. *Nature* **332**, 272–275 (1988).
10. Lehmann, J. M., Riethmüller, G. & Johnson, J. P. *Nucleic Acids Res.* **18**, 1048 (1990).
11. Otsu, M. *et al.* *Cell* **65**, 91–104 (1991).
12. Aroian, R. V., Koga, M., Mendel, J. E., Ohshima, Y. & Sternberg, P. W. *Nature* **348**, 693–699 (1990).
13. Han, M. & Sternberg, P. W. *Cell* **63**, 921–931 (1990).

to know how critically their solution depends on the assumed values of some of the physical properties of the Earth's core and on the basic assumptions of their method.

Buffett *et al.* conclude that compositional and thermal buoyancy fluxes in the outer core are of comparable magnitude. In this regard, Stevenson *et al.*⁷ suggested that the mode of powering the geodynamo may have changed during geological time. In the Earth's early history, the magnetic field was generated by thermal convection. After nucleation of the inner core (which they estimate to have happened 1,500–2,500 million years ago), the release of gravitational energy

ASTRONOMY

became the dominant source. It is interesting to speculate that the reason that Mars and Venus have essentially no magnetic field is that they do not have an inner core. □

J. A. Jacobs is in the Institute of Earth Studies, The University College of Wales, Aberystwyth, Dyfed SY23 3DB, UK.

1. Buffett, B. A., Huppert, H. E., Lister, J. R. & Woods, A. W. *Nature* **356**, 329–331 (1992).
2. Lehmann, I. *Bur. Centr. Seismol. Int.* **A14**, 3–31 (1936).
3. Oldham, R. D. Q. *Jl geol. Soc. Lon.* **62**, 456 (1906).
4. Birch, F. *Am. J. Sci.* **238**, 192–211 (1940).
5. Kelvin, Lord J. *Victoria Inst. London* **31**, 11–35 (1899).
6. Gubbins, D., Masters, T. G. & Jacobs, J. A. *Geophys. J. R. astr. Soc.* **59**, 57–99 (1979).
7. Stevenson, D. J., Spohn, T. & Schubert, G. *Icarus* **54**, 466–489 (1983).

Convective motion on the Sun

Nigel Weiss

HIGH-resolution observations are revealing the pattern of convection at the surface of the Sun. On page 322 of this issue¹, Muller *et al.* describe white-light observations made during an interval of exceptionally clear seeing at the Pic du Midi Observatory in the French Pyrenees. For a period of 3 hours they obtained images that were virtually unaffected by atmospheric distortion, so that resolution was limited only by the aperture of the telescope. This is the best sequence so far obtained and, after digital processing at the Lockheed Palo Alto Research Laboratory, it provides fascinating details about the structure of solar convection.

Seen in white light, the solar photosphere shows the well-known granulation first observed by William Herschel almost 200 years ago. Individual bright granules are surrounded by a network of cool, dark material and are separated by distances of 1,000–2,000 kilometres; the hot, bright material is rising and the cool material is sinking, so these cells are evidence of convection, which carries energy up to just below the visible surface². By following the proper motions of individual granules, larger-scale cellular patterns (mesogranules and supergranules) can be detected. The new observations show how these structures evolve.

For the first time it is possible to explore the relationship between them, which provides clues to understanding the structure of convection deep below the surface. Explaining these velocity patterns is important not only in itself but also because of their interactions with solar oscillations and with magnetic fields. Although the Sun is the only star where such motions can be observed in detail, similar behaviour must occur in other cool stars. Moreover, the Sun is a

unique laboratory for studying turbulent magnetoconvection in physical conditions that are unobtainable on Earth.

Techniques for processing such data were first developed at Lockheed in connection with the Spacelab-2 mission in 1985, but the data sets obtained then were limited to 28 minutes by the orbital period of the satellite. After aligning and 'destretching' the images it is necessary to eliminate distortion caused by solar oscillations with periods around 5 minutes. This is done by Fourier-analysing the data in two space dimensions and time, and then filtering out all super-sonic variations. The resulting images are so clear that it is possible to track the proper motions of granules and hence to determine large-scale velocity patterns^{3,4}. These methods have firmly established the existence of three different non-overlapping scales of motion: in addition to granules there are mesogranules, with a characteristic diameter of 3,000–10,000 kilometres, as well as the supergranules, with diameters around 30,000 kilometres, which are outlined by weak magnetic fields and have been known for 30 years.

Measurements of mesogranular velocities are complicated by the presence of exploding granules. These are particularly vigorous granules, which expand rapidly, displacing their neighbours, until they develop a cool core and break up. The Spacelab-2 observations showed that exploders occur preferentially near the centres of mesogranules and raised the possibility that mesogranular motion might be just the cumulative effect of these exploders⁴. Careful inspection of the new data from the Pic du Midi indicates, however, that individual exploders move in a systematic mesogranular velocity field and that mesogranules survive for several hours, much longer

than the lifetimes of exploders⁵. Apparently both structures contribute to the measured motion and it seems likely that they are symbiotically linked⁶.

Even more interesting is the relationship between mesogranules and the long-lived supergranules. The 3-hour run shows that mesogranules appear near the centre of a supergranule and are transported towards its boundaries, where they are destroyed. This is the most important result in the new paper. How should it be explained? One possibility is that these patterns represent two distinct scales of motion; another, perhaps more plausible, is that mesogranular motion is caused by parasitic instabilities of thermal boundary layers in supergranular convection cells⁶.

What can be learnt from these observations about the structure of convection below the visible surface? Numerical experiments confirm that the stratification favours broad upwellings surrounded by rapidly sinking sheets which are focused into isolated plumes^{7–9}. One possibility is that the plumes merge to give a continuously varying self-similar structure⁷. Another, suggested by the surface observations, is that there are coherent structures with separate scales. In that case, one might hope to detect structures with diameters comparable with the depth (200,000 kilometres) of the convection zone — but such giant cells have not been found. It now seems more likely that isolated supergranular plumes can sink almost to the base of the convection zone⁶.

It is essential to provide a proper theory of stellar convection which is able to describe the velocity structures observed at the surface of the Sun. Only then will it be possible to explain the differential rotation profiles derived from helioseismology. If that is achieved, the next goal will be to produce a realistic model of the solar dynamo, which is responsible for magnetic activity in the solar atmosphere, including sunspots¹⁰ and flares¹¹, which are being observed with ever-increasing precision. □

Nigel Weiss is in the Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge CB3 9EW UK.

1. Muller, R. *et al.* *Nature* **356**, 322–325 (1992).
2. Stix, M. *The Sun* (Springer, Berlin, 1989).
3. Simon, G. W. *et al.* *Astrophys. J.* **327**, 964–967 (1988).
4. Title, A. M. *et al.* *Astrophys. J.* **336**, 475–494 (1989).
5. Simon, G. W., Title, A. M. & Weiss, N. O. *Astrophys. J.* **375**, 775–788 (1991).
6. Simon, G. W. & Weiss, N. O. *Mon. Not. R. astr. Soc.* **252**, 1p–5p (1991).
7. Stein, R. F. & Nordlund, A. *Astrophys. J.* **342**, L95–L98 (1989).
8. Spruit, H. C., Nordlund, A. & Title, A. M. *Rev. Astr. Astrophys.* **28**, 263–301 (1990).
9. Cattaneo, F. *et al.* *Astrophys. J.* **370**, 282–294 (1991).
10. Livingston, W. *Nature* **350**, 45–46 (1991).
11. Golub, L. *et al.* *Nature* **344**, 842–844 (1990).

Gone fishing

THE changing ratios of strontium isotopes in sea water can be used to establish the chronology of marine sediments. But an ingenious scheme to correlate this record with continental sediments proved too slippery to handle (P. L. Koch *et al.* *Earth planet. Sci. Lett.* **108**, 227–287; 1992). Strontium replaces calcium in vertebrate bone, and if samples of bone formed in the oceans could be found in continental sediments, this bone could help correlate the disjunct land and ocean geological records. Koch *et al.* came up with the very thing — 15-million-year-old migratory salmon. Strontium isotope ratios from the continental fossils of mature salmon should reflect those of their adopted marine homes. But sad to say, it turned out that the vagaries of diagenesis mean that strontium ratios in fossils reflect the sediments in which they are found rather than the oceans once inhabited by the fish.

Getting the point

A CIRCLE of diagnostic understanding of Duchenne muscular dystrophy has been closed by R.G. Roberts *et al.* (*Proc. natn. Acad. Sci. U.S.A.* **89**, 2331–2335; 1992). About two-thirds of the mutations to the dystrophin gene, the underlying cause of muscular dystrophy, are large-scale rearrangements which can be directly characterized and used in screening; the remainder, however, have proved hard to identify because the gene is so large and complex. Roberts *et al.* analysed the coding sequence of genes from seven patients showing no gross defects. In each they identified a point mutation which would cause a premature termination of transcription and be enough to account for the disease.

Any which way

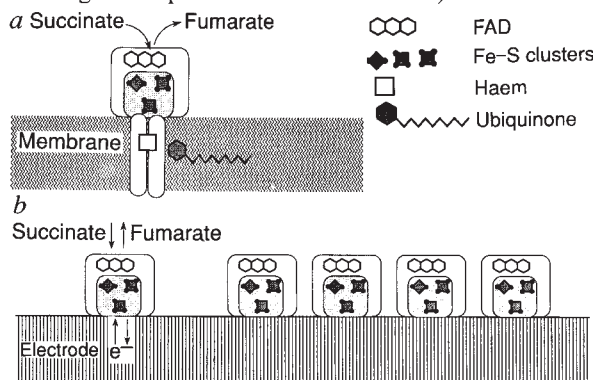
THE source of astronomical γ -ray bursts remains a puzzle, despite the furious rate at which the satellite-borne Compton Observatory is identifying them. Only in January, operators reported an analysis of the first 153 bursts detected. Now C.A. Meegan *et al.* (*IAU Circ.* 5478; 1992) describe over 100 more. But still the message is ambiguous. The bursts are distributed isotropically over the sky, and because our Galaxy is a flattened disk, that suggests the bursts originate beyond its confines, in other galaxies. But there seems to be a paucity of the faintest (most distant) bursts, implying that they are more common nearby than far away. Nor is an excess of bursts evident when looking towards the Large Magellanic Cloud or towards Andromeda, our galactic neighbours. With the errors in the burster distribution growing as data accumulate, there is plenty of scope for inventive solutions.

Plug in a molecular diode

Richard Cammack

ON page 361 of this issue¹ Sucheta and colleagues describe a remarkable enzymatic reaction which appears to behave like a diode; this property may turn out to be of physiological significance.

The protein concerned is succinate dehydrogenase, a mitochondrial enzyme that catalyses the oxidation of succinate to fumarate^{2,3}. Redox reactions of proteins can be observed by dynamic electrochemical methods, but such methods usually require the addition of compounds known as promoters (to assist binding of the proteins to the electrodes)



a, Reaction of succinate-ubiquinone reductase in a lipid bilayer membrane. b, Proposed arrangement of succinate dehydrogenase dimers as a monolayer on a graphite electrode surface.

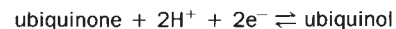
or mediators (to assist electron transfer)^{4,5}. Sucheta *et al.* demonstrate that succinate dehydrogenase can bind to the surface of a graphite electrode, and can display rapid electron transfer without assistance. Unexpectedly, the enzyme prefers to conduct electrons only one way — it catalyses the oxidation of succinate to fumarate very well, but when a strong reducing potential is applied it catalyses the reverse pathway only poorly. It acts, as the title of the paper declares, like a molecular diode. This effect may be of physiological importance in helping to prevent the Krebs cycle from back-pedalling, and it may also point the way to the development of new types of bio-electronic device.

Succinate dehydrogenase is the only membrane-bound enzyme of the Krebs citric acid cycle. It forms part of a respiratory complex, succinate-ubiquinone reductase. The final oxidant for succinate is not, as students used to be taught, the flavin FAD, but a membrane-bound carrier, ubiquinone⁶. To carry out this apparently straightforward transfer of two hydrogen atoms, the enzyme has a complex electron-transfer system, consisting of FAD, three iron-sulphur clusters and a *b*-type cytochrome (see *a* in the figure). The enzyme comprises two hydrophilic protein subunits,

which react with the soluble succinate:



The electrons are then fed through the protein to the two hydrophobic subunits in the membrane, where they recombine with two hydrogen ions and ubiquinone:



Armstrong, Hill and others^{4,5} have adapted the conventional techniques of voltammetry to measure redox potentials and redox reactions of the proteins, making it possible to study the kinetics of electron transfer in the millisecond time range. The method has also been used to follow the rapid interconversions of iron-sulphur clusters of the [3Fe-4S] and [4Fe-4S] types within proteins⁷, and even the insertion of unusual metal ions such as thallium⁸.

With succinate dehydrogenase, Sucheta *et al.* found that a graphite electrode presents a particularly receptive surface of hydrophobic patches and negatively charged carboxylate groups. The enzyme readily catalysed the oxidation of succinate to fumarate, giving electrons to the electrode surface, just as it does in the mitochondrial membrane. But when the enzyme's capability to catalyse the reverse reaction was addressed, fumarate reduction proceeded rapidly only under conditions of small driving force. Application of a more negative potential (which from classical considerations should accelerate the reduction) resulted instead in a dramatic attenuation of catalytic activity. This atypical behaviour contrasts with that of fumarate reductase, a closely related enzyme, which in anaerobic bacteria catalyses the reverse reaction from fumarate to succinate, and which shows reversible redox chemistry at the electrode (ref. 5 and F. A. Armstrong, personal communication). The intriguing behaviour of succinate dehydrogenase seems to be an intrinsic property of the enzyme, and is consistent with observations from steady-state kinetics of the enzyme in solution^{1,9}.

A tendency towards unidirectional electron transfer is also evident in other electron-transfer proteins, such as the hydrogenases, which catalyse the production or consumption of hydrogen gas by microorganisms. Some of these en-

zymes are most effective in the direction of hydrogen production, and others in hydrogen uptake, even with the same electron carriers¹⁰.

Proteins do not have the solid-state properties of semiconductors. The diode-like behaviour in this case is a consequence of the microscopic electron-transfer reactions within each individual protein molecule. Sucheta *et al.* suggest that a kinetic mechanism may be involved, in which the binding and dissociation of the substrates to the enzyme are hindered unless the flavin is oxidized. Under strong reducing conditions the flavin is reduced, so the reaction is inhibited. Other mechanisms are possible, for example involving one-electron-reduced semiquinone states of the flavin.

What the new results mean in physiological terms is unclear, though it may be that the 'ratchet-like' capacity prevents the enzyme from running backwards under conditions where the mitochondria are anaerobic, for example in muscle during intensive exercise. From the molecular electronic point of view, the diode-like behaviour of enzymes is of interest because enzymes are well able to change their catalytic properties in response to effector molecules, such as inhibitors and activators, or to other macromolecules. If the reverse flow of electrons through an enzyme could be made conditional on such an effector molecule, the resulting 'device' would have the characteristics of a molecular transistor. That prospect may be only a gleam in the eye, but it is an enticing one. □

Richard Cammack is in the Division of Biomolecular Sciences, King's College London, Campden Hill Road, London W8 7AH, UK.

1. Sucheta, A., Ackrell, B. A. C., Cochran, B. & Armstrong, F. A. *Nature* **356**, 361–362 (1992).
2. Ackrell, B. A. C., Johnson, M. K., Gunsalus, R. P. & Cecchini, G. in *Chemistry and Biochemistry of Flavoenzymes* Vol. 3 (ed. Mueller, F.) 229–297 (CRC, Boca Raton, Florida, 1992).
3. Ohnishi, T. in *Current Topics in Bioenergetics* Vol. 15 (ed. Lee, C. P.) 37–65 (Academic, New York, 1987).
4. Frew, J. E. & Hill, H. A. O. *Eur. J. Biochem.* **172**, 261–269 (1988).
5. Armstrong, F. *Struct. Bond.* **72**, 137–221 (1990).
6. Cammack, R. *Trends biochem. Sci.* **12**, 377 (1987).
7. Armstrong, F. A., George, S. J., Cammack, R., Hatchikian, E. C. & Thomson, A. J. *Biochem. J.* **264**, 265–273 (1989).
8. Butt, J. N. *et al. J. Am. chem. Soc.* **113**, 8948–8950 (1991).
9. Zejilemaker, W. P., Dervartanian, D. V., Veeger, C. & Slater, E. C. *Biochim. biophys. Acta* **178**, 213–224 (1968).
10. Adams, M. W. W. *Biochim. biophys. Acta* **1020**, 115–145 (1990).

Hydrogen atoms band together

Risto Nieminen

ELECTRONS moving through a crystal have their energies divided up into different bands, separated by band gaps. Whether a material is insulating, semiconducting, semimetallic or metallic depends on the nature of these bands, on the band-gap separation between them and on the way they are populated. The bands and gaps arise because the travelling electrons are diffracted by the crystal lattice, and transmission is completely impossible for electrons of certain energies. What is good enough for electrons is good enough for protons, their

chemical counterparts 2,000 times more massive, report C. Astaldi *et al.*¹, who have discovered band-like behaviour of hydrogen atoms moving over the surface of copper.

It is quite natural to think of situations, other than with electrons, in which the wave nature of a particle would have similar consequences — not only diffractive effects (as with X-rays and cold neutrons), but also a clear band structure. As discussed recently by Pendry², periodic arrays of holes drilled into a dielectric material lead to a band

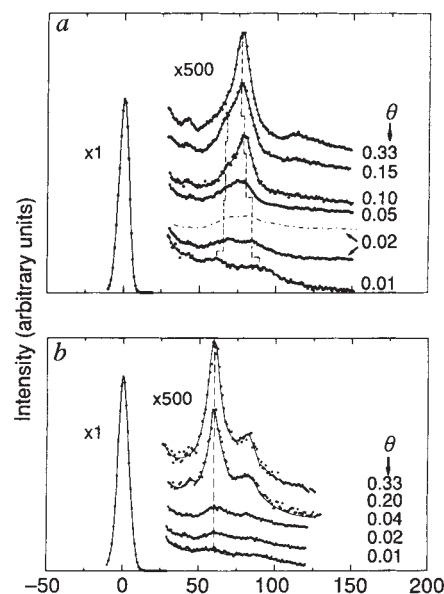


FIG. 2 High-resolution electron-energy-loss spectra¹ of the hydrogen isotopes H (a) and D (b) chemisorbed on a Cu(110) surface, measured for different values of the relative coverage θ . The broadening and line shape reveal the delocalization of H.

structure for photons. In particular, it is possible to make structures where electromagnetic waves of certain wavelengths do not propagate at all: they are 'photonic insulators'. Defects purposely introduced into the periodic structure lead to the appearance of 'impurity states', in complete analogy with acceptor-doped semiconductors and insulators.

The new experiments of Astaldi *et al.* take these studies in another direction. The authors use high-resolution electron-energy-loss spectroscopy to investigate the disposition of hydrogen atoms on a copper surface. In this type of experiment, electrons with an energy of a few electronvolts are reflected off the coated surface, and their recoil energy is analysed with a resolution of a few millielectronvolts. The electrons can transfer part of their energy to the surface atoms which may, for example, start

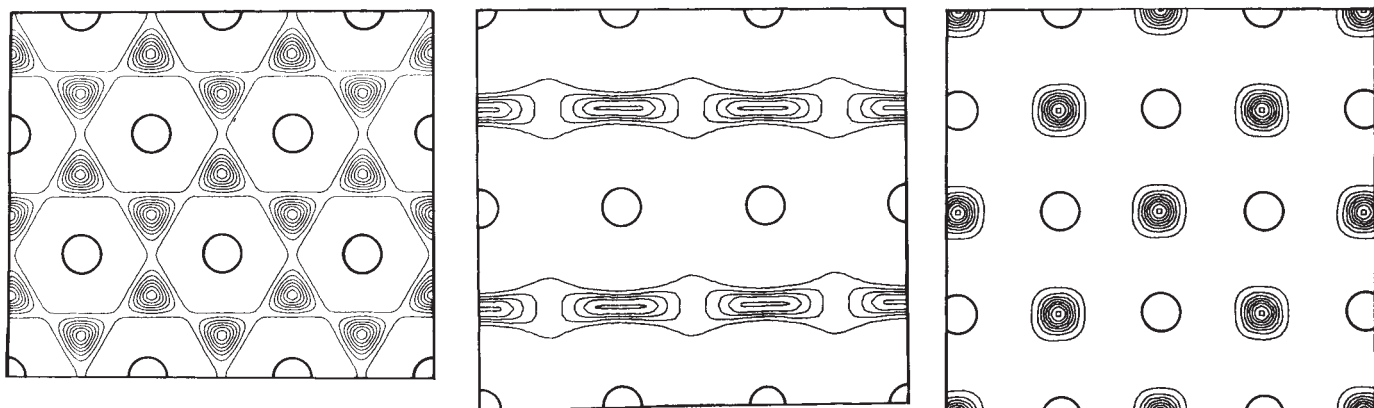
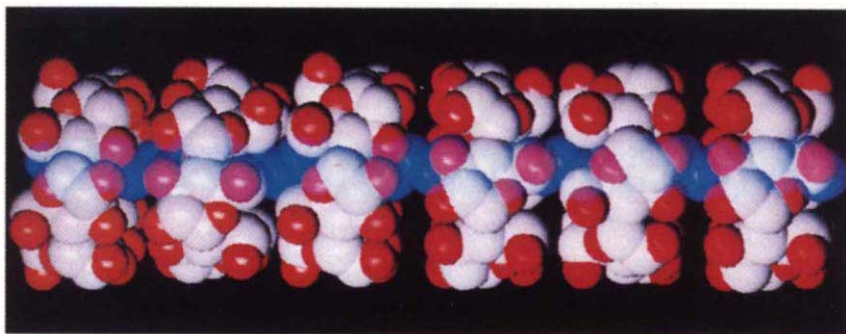


FIG. 1 Ground-state wavefunctions for hydrogen chemisorbed at the three low-index surfaces of face-centred cubic nickel. From left to right, (100), (110) and (111). Note the delocalization of the wavefunction on the (110) surface.

Picking up the chemical thread



THREADING the molecular needle — passing a linear molecule through a cyclic one — is an old trick: 'hooplane', an alicyclic ketol on a hydrocarbon axle, was fashioned in 1967 by researchers at Syntex Research in Palo Alto, and the term 'rotaxane' for these rotor-and-axle assemblies dates back to Gottfried Schill's similar construction in 1980. But whereas these were the products of conventional ring-closure syntheses, the principles of supramolecular host-guest chemistry are now making the molecular gymnastics of rotaxane synthesis far easier. The extent to which this is true should be apparent from the supramolecular complex described by A. Harada *et al.* on page 325 and pictured above, which contains twenty or more cyclodextrin rotors on a strand of polyethylene glycol.

A wide variety of threaded supermolecules has been developed in recent years, many by Fraser Stoddart's group and Jean-Pierre Sauvage and colleagues. Stoddart has used linear and cyclic paraquat-hydroquinone complexes to create the 'molecular shuttle', a rotaxane in which the rotor can switch between two docking points on the thread, and a range of catenanes, in which the ends of the thread are joined. A single bead threaded on a long loop with several docking points constitutes a 'molecular train' that can circle around the line until cooled to a standstill. Stoddart has made the whimsical suggestion that several beads on a strand might constitute a 'molecular abacus'.

The complex synthesized by Harada *et al.* shows that this is easier than might be supposed. The components are simply mixed in aqueous solution, and the resulting complex is capped at each end with 2,4-dinitrofluorobenzene. Diffraction and spectroscopic studies of the crystalline form confirm that it is a 'molecular necklace'. The polymer thread is about 266 Å long, and the cyclodextrins occupy only about 160 Å or so of this length — Harada *et al.* suspect that each sticks closely to its neighbours via hydrogen bonding when in the solid state, but that the molecular beads are mobile when the complex is in solution.

P.B.

to vibrate around their surface sites. Energy and momentum conservation laws, together with the symmetries of the ground and excited states, yield distinct selection rules for what transitions are possible. The loss spectra and their dependence on scattering angles and adatom coverage can be used to obtain details on the nature of the adsorbed particles.

A simple measure of the importance of quantum effects for adsorbates can be obtained by comparing two competing effects. On the one hand, the more strongly the adsorbate prefers a particular site on the surface, the greater the potential-energy barrier separating that site from its neighbouring equivalents. On the other hand, the more tightly confined an adsorbate is, the greater is its quantum-mechanical zero-point motion about its equilibrium position. In the case of conduction electrons in a crystal, the zero-point energy overwhelms the binding energy, and the electrons are free to move throughout the material.

Cole *et al.* pointed out several years ago³ that in the case of physisorbed (weakly bound) helium on graphite, the zero-point energy is greater than the potential barriers and the resulting film can assume quantum effects. Although hydrogen is much more strongly bound on metals (chemisorbed), its mass is also four times smaller, so that its zero-point energy is raised proportionately. So quantum behaviour is not totally unexpected in this case. In fact, hints of unusual (that is, nonclassical) properties have accrued over the years.

Electron-diffraction experiments⁴ have revealed substantial disorder for low-coverage hydrogen phases, indicative of large adatom mobility. Surface diffusion monitored by field-emission studies⁵, shows anomalous behaviour suggestive of large tunnelling effects. Electron-energy-loss spectra and ion-scattering experiments were also difficult to reconcile with a picture of a classical particle vibrating around its equilibrium position on the surface.

Intrigued by these results, my col-

leagues and I performed quantitative calculations for the proton band structure on various nickel surfaces, discussing its consequences for diverse physical phenomena^{6,7}. These calculations, which allow for the screening of the proton charge by the metallic electrons, show substantial delocalization of the adatoms (Fig. 1, preceding page). In particular, we pointed out that the proper analysis of electron energy-loss spectra should be made in terms of the excitation of protons between different extended bands, not in terms of the vibrations of atoms tied to individual sites on the surface.

Astaldi *et al.* performed their experiments on the (110) face of copper. For this surface, expected to be qualitatively similar to Ni(110), the proton potential is such that hydrogen moves mainly along essentially one-dimensional troughs in the [110] direction. The spectra (Fig. 2) were taken for electrons scattering specularly in this direction, with the primary energy of 2.5 eV. The loss peaks at 79 and 118 meV correspond to dipole-allowed excitations of the surface hydrogens. The important observation is the peculiar coverage- and isotope-dependence of the spectra, which provide spectroscopic evidence for the gradual transition from quantum delocalized behaviour at low coverages to localized, ordered adatom structures at high coverages. With increasing hydrogen coverage, the mobile 'liquid' of adsorbed protons forms a highly correlated system with very strong short-range repulsive interactions. Astaldi *et al.* show that their experiments are consistent with the quantum picture by analysing it in terms of a one-dimensional model hamiltonian. The peak structure as well as the decrease of the linewidth with coverage can be explained.

The quantum motion of hydrogen on surfaces can have interesting consequences for various processes. There is already evidence for nonthermal mobility of hydrogen at low temperatures, both in bulk crystals and on surfaces. The proton motion has a component of quantum-mechanical tunnelling through potential barriers instead of mere classical hopping. This in turn may influence such surface reactions (as in catalysis) where hydrogen diffusion is an important reaction step: the adatoms can be hot in the sense that their mobility is much larger than one might expect with simple thermal motion.

1. Astaldi, C. *et al.* *Phys. Rev. Lett.* **68**, 90–93 (1992).
2. Pendry, J. *Nature* **354**, 435–436 (1991).
3. Cole, M. W. *et al.* *Rev. mod. Phys.* **53**, 199–245 (1981).
4. Christmann, K. *et al.* *J. chem. Phys.* **70**, 4168–4180 (1979).
5. DiFoggio, R. & Gomer, R. *Phys. Rev.* **B25**, 3490–3525 (1982).
6. Puska, M. J. *et al.* *Phys. Rev. Lett.* **51**, 1083–1087 (1983).
7. Puska, M. J. & Nieminen, R. M. *Surf. Sci.* **157**, 413–435 (1985).

Surface hydrogen is thus a candidate for an interesting, strongly correlated quantum liquid, intimately coupled to the underlying surface and its excitations. A number of hydrogen-related properties have been intensely studied by surface physicists. The competition between hydrogen-substrate and hydrogen-hydrogen interactions leads to a wealth of ordering phenomena of both hydrogen adatoms (surface phases) and substrate atoms (surface relaxation and reconstruction). The new quantum aspect discussed here is the importance

of zero-point motion. Yet to be explored are the possible consequences of quantum statistics for hydrogen-related surface phenomena. Should one consider screened hydrogen adatoms moving around on a metal surface as fermions (like electrons), or as bosons, like the Cooper pairs of electrons responsible for superconductivity? Can the system exhibit superfluidity? □

Risto Nieminen is in the Laboratory of Physics, Helsinki University of Technology, 02150 Espoo, Finland.

CHAGAS' DISEASE

Defence cuts begin to bite

Peter Parham

In April and June of 1990, two groups reported the creation of mice lacking class I major histocompatibility complex (MHC) molecules^{1,2}. This they accomplished by knocking out the gene encoding β_2 -microglobulin (β_2m), the ubiquitous light chain of class I molecules. Because of this disruption, β_2m -negative mice lack CD8⁺ T cells, lymphocytes which depend upon class I molecules for their very existence. In other respects, however, the mice seemed to be normal and, contrary to expectation, were neither sickly nor difficult to breed. As cytotoxic CD8⁺ T cells are believed to defend against viral infections, it was thought that β_2m -negative mice would readily succumb to viruses³, and with this in mind immunologists have been trying to 'take out' the β_2m -negative mouse since the summer of 1990. So far, traditional strategies of viral attack have been evaded⁴, but success has been accomplished with a smarter weapon from the arsenal. On page 338 of this issue⁵ Tarleton *et al.* report that infection with the American trypanosome, *Trypanosoma cruzi*, always results in the death of β_2m -negative mice.

Trypanosoma cruzi causes Chagas' disease, a virtually incurable and often lethal condition which affects millions of people in Latin America⁶⁻⁸. This protozoan infects most mammalian cells, and numerous wild and domestic species provide reservoirs for the parasite in endemic areas. The life cycle runs through three morphologically distinct forms: a blood-borne trypomastigote that infects mammalian cells; an intracellular amastigote that replicates therein; and an epimastigote that lives in the gut of insects that transmit the disease (see figure). On being bitten, "parasitic inoculation occurs by scratching the bug faeces into the insect bite wound, or by chance carry over of the contaminated faeces to intact mucosa or conjunctiva"⁶.

For adults the resulting parasitic infec-

tion is usually benign and often goes unnoticed; for children it more frequently produces an acute Chagas' disease in which parasites can readily be detected in the blood and other tissues. Host defences eventually reduce the parasite load and the disease subsides after one

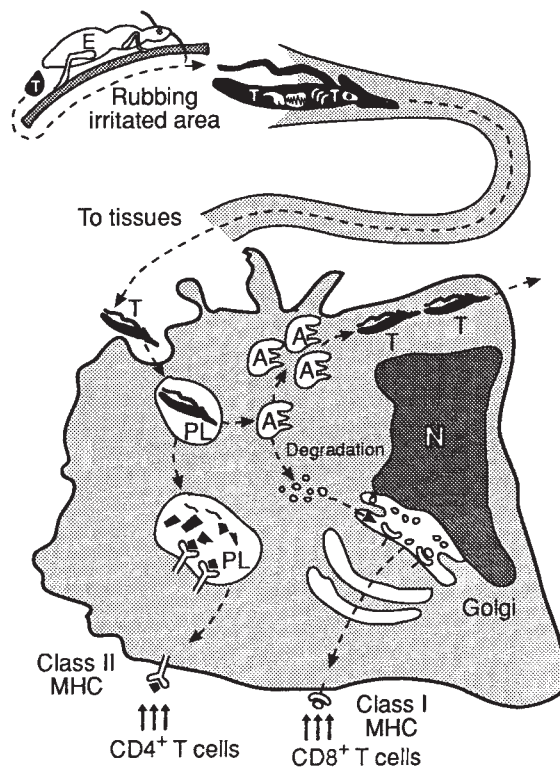
or two months. But for certain individuals, infection starts a clock running and after latency periods that can last for decades, the chronic form of Chagas' disease erupts. It is characterized by inflammatory reactions in muscle and lack of detectable parasites, and frequently causes death from heart failure⁶⁻⁸.

Chagas' disease has characteristics associated with autoimmunity. An immediate threat — in this case parasite infection — is contained; but in the course of mounting defence, lymphocytes that crossreact with self-antigens of healthy cells increase in numbers. Following acute infection, the continued activities of these autoimmune cells lead to inexorably accumulating tissue damage and eventual pathology. Debate as to whether chronic Chagas' is an autoimmune disease has run for decades (for review see refs 6,9,10) and led to the demonstration of autoimmune reactions in humans and animals infected with *T. cruzi*. Furthermore, an array of cross-reactive antigens shared by *T. cruzi* and mammalian cells, especially in nervous tissue, has been described⁹. The outcome of the debate has

practical consequences, for if chronic Chagas' is caused directly by lymphocytes and not parasites, then vaccination becomes a less attractive strategy for protection (the concern here being that vaccination would actually induce the disease-causing immune response). For this reason, claims that chronic Chagas' is an autoimmune disease have been viewed with a very critical eye¹⁰.

Laboratory mice infected with *T. cruzi* suffer a course of acute and chronic infection having many similarities to the human disease. Investigation of this experimental Chagas' disease is directed to dissection of the immune response to *T. cruzi* and resolution of the role of autoimmunity. Infection with *T. cruzi* turns out to be a heavy-handed affair in which the host's defences go badly awry. Massive nonspecific polyclonal activation of B and T cells is induced and a state of general immunosuppression ensues. In time, normal mice resist the invasion and their recovery, as well as the chronic condition, is believed to depend upon CD4⁺ T cells¹¹⁻¹³.

While others have been



Aspects of the life cycle of *Trypanosoma cruzi* and the potential for antigen presentation to CD4⁺ and CD8⁺ T cells (E, epimastigotes, which multiply in insect vectors; T, blood-borne trypomastigotes, which enter tissue macrophages by phagocytosis). In the phagolysosomes (PL), trypomastigotes, by a mechanism that is unclear, enter the cytoplasm and convert to replicating amastigotes (A) which eventually produce new trypomastigotes. In the phagolysosome system there is opportunity for trypanosome antigens to be presented by class II MHC molecules to CD4⁺ T cells; during replication in the cytoplasm there is the possibility for trypanosome antigens to be presented by class I MHC molecules. N, nucleus.

investing time in study of CD4⁺ cells, however, Tarleton has been building a case for the involvement of CD8⁺ T cells. Previously, he showed that depletion *in vivo* of CD8⁺ T cells with anti-CD8 monoclonal antibody rendered mice susceptible to *T. cruzi*, the infected mice dying within a month. CD8⁺ T cells seemed to contribute to the establishment of protective immunity but not its long-term maintenance, because depletion of CD8⁺ cells from chronically infected mice compromised neither control of the infection nor the capacity to resist reinfection¹⁴.

The β_2m -negative mouse lacks CD8⁺ T cells and thus provided Tarleton *et al.*⁵ with a new system to assess their importance for defence against *T. cruzi*. The experiment was simple — infect animals with *T. cruzi* and see if they die — and the results were unequivocal: the β_2m -negative mice all die whereas their heterozygous and wildtype litter mates survive. As with other experiments on transgenic or 'knocked-out' mice, the interpretation is not so clear cut. It is safe to conclude that β_2m or β_2m -associated class I molecules contribute to resisting *T. cruzi*, but to invoke CD8⁺ T cells means assuming that the only effect of β_2m is in the selection and function of CD8⁺ T cells. This is a brave premise, and one that Tarleton *et al.* acknowledge is a simplification. In particular, class I MHC molecules affect the specificity and function of natural killer (NK) cells as well as CD8⁺ cells¹⁵, and NK cell development is compromised in β_2m -negative mice^{16,17}. Natural killer cells are important during the early phases of many infections; as a good number of them express the CD8 molecule, one can make a case that the results of Tarleton and coworkers^{5,14} are consistent with resistance to *T. cruzi* depending upon NK cells, as has already been suggested¹¹.

Resolution of this and other possible explanations will require attempted rescue of β_2m -negative mice from *T. cruzi* attack with infusions of defined cell populations. That CD8⁺ cytotoxic T cells contribute to defence against *T. cruzi* is certainly plausible, given the parasite's life cycle. Amastigote replication takes place within the host cell's cytoplasm, the precise cellular compartment from which peptides presented by class I MHC molecules are commonly derived (see figure). Such presentation of *T. cruzi* antigens could stimulate CD8⁺ cytotoxic T cells, leading to destruction of infected cells and interference with parasite replication.

Contrasting with the finesse with which the β_2m -negative mouse disposed of influenza virus⁴, it seems to have just given up in the face of *T. cruzi*. On the evidence of Tarleton *et al.*, the mice

became completely overrun with parasites before their defences could fire off a significant shot. Immune activities seen in normal mice seem to be attenuated or running at slower tempo in their β_2m -negative siblings: the induced immunosuppression was less, parasite numbers were higher and acute inflammatory reactions in muscular tissues much reduced. The incontrovertible message from Tarleton *et al.* is that, despite a façade of laboratory vigour, β_2m -negative mice collapse on confrontation with a hardened Third World parasite.

Despite success in identification of components of the immune system, immunologists are still a long way from a precise description of events following trespass of an infectious organism on mammalian territory. The black box standing between insect bite and heart failure in Chagas' has its equivalent in every infectious and autoimmune disease. Since the first description of the β_2m -negative strains, mice knocked out for an increasing number of immune components have been produced and the general result is viable, healthy, fecund mice. With judicious breeding of these mice, the immune system can gradually be pared away until the 'minimal mouse' — lacking immunoglobulins, T-cell receptors, MHC molecules, co-receptors, lymphokines and all the other weaponry of immunity — is left. The availability of such mice will create a new order in the analysis of host-pathogen relationships, in that they will provide a system in which the immune response can be reconstituted using purified proteins, cloned cells and one's chosen pathogen. □

Peter Parham is in the Departments of Cell Biology, and Microbiology and Immunology, Stanford University School of Medicine, Stanford, California 94305, USA.

1. Zijlstra, M. *et al.* *Nature* **344**, 742–746 (1990).
2. Koller, B. H., Marrack, P., Kappler, J. W. & Smithies, O. *Science* **248**, 1227–1230 (1990).
3. Parham, P. *Nature* **344**, 709–711 (1990).
4. Eichelberger, M., Allan, W., Zijlstra, M., Jaenisch, R. & Doherty, P. C. *J. exp. Med.* **174**, 875–880 (1991).
5. Tarleton, R. L., Koller, B. H., Latour, A. & Postan, M. *Nature* **356**, 338–340 (1992).
6. Santos-Buch, C. A. *Int. Rev. exp. Path.* **19**, 63–100 (1979).
7. Pereira, M. E. A. in *Modern Parasite Biology* (ed. Wyler, D.) 64–78 (Freeman, New York, 1990).
8. Brener, Z. & Krettl, A. U. in *Modern Parasite Biology* (ed. Wyler, D.) 247–261 (Freeman, New York, 1990).
9. Petry, K. & Eisen, H. *Parasitology Today* **5**, 111–116 (1989).
10. Kierszenbaum, F. *J. Parasitol.* **72**, 201–211 (1986).
11. Rottenberg, M., Cardoni, R. L., Andersson, R., Segura E. L. & Orn, A. *Scand. J. Immun.* **28**, 573–582 (1988).
12. Younes-Chennoufi, A. B., Said, G., Eisen, H., Durand, A. & Hontebeyrie-Joskowicz, M. *Trans. R. Soc. trop. Med. Hyg.* **82**, 84–89 (1988).
13. Ribeiro-dos-Santos, R., Pirmez, C. & Savino, W. *Res. Immun.* **142**, 134–137 (1991).
14. Tarleton, R. L. *J. Immun.* **144**, 717–724 (1990).
15. Storkus, W. J. & Dawson, J. R. *CRC Crit. Rev. Immun.* **10**, 393–416 (1991).
16. Liao, N.-S., Blix, M., Zijlstra, M., Jaenisch, R. & Raulet, D. *Science* **253**, 199–202 (1991).
17. Hoglund, P. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **88**, 10332–10336 (1991).

Pinched diamonds

DIAMOND is the most tantalizing of materials. It is uniquely strong, stiff, thermally stable and thermally conductive, and is transparent over an amazingly wide spectral bandwidth. Yet to synthesize it from graphite is absurdly difficult, usually needing clever catalysts and desperate extremes of heat and pressure. A recent synthesis has formed diamond as tiny fibres. Daedalus now plans to make diamond fibre by the kilometre.

His process exploits the pinch effect, which compresses a current-carrying conductor in its own magnetic field. A large current passed briefly through a catalyst-doped carbon fibre should heat it so strongly, and pinch it to such a pressure, that it would collapse to diamond. For a fibre 0.01 mm across, a mere five thousand amps should do the trick. Carbon fibre passed through two sets of rollers with a suitable potential between them could be continuously 'pinched' into diamond fibre, and wound on a receiving drum. The transformation is martensitic and very rapid; so the process could be run fast enough to make diamond fibre a bulk commodity.

Diamond fibre will be a wonderful new material. Plastics reinforced with it will be many times stronger and stiffer than rival composite materials. On diamond-fibre cables, suspension bridges and power lines will soar many kilometres in a single span. Wide-band diamond optical fibres will carry thousands of trashy TV channels simultaneously. On a more intimate scale, it will be woven into marvellous silky lingerie, sheer, lustrous, utterly glamorous — and also snag-resistant, stain-rejecting, indigestible to moths, impervious to heat, and bullet-proof. Indeed, diamond-fibre fabrics will revolutionize the design of flak-jackets, military uniforms, fire-blankets and protective clothing generally.

Variants on the basic product will also be possible. By running fine carbon tube through the pinch-mill, diamond hypodermic tubing should result. Similarly, a carbon-coated wire feedstock would give diamond-plated wire. With its marvellous mechanical strength, electrical insulation and unprecedented heat-transfer properties, it should be widely welcomed by the electrical and electronics industries. Furthermore, diamond-pinching could easily be scaled up. A few million amps could pinch a thick cylindrical graphite feedstock into diamond rod several centimetres across. It could be made into all sorts of novel diamond lenses, crucibles and cutting tools. But probably not jewellery. Bulk pinched diamond will be so cheap that nobody will want to advertise their poverty by wearing the stuff.

David Jones

Measuring shelf afterlife

SIR — The radioactivity of substances in daily use has been well documented. But it is not common knowledge even to nuclear and health physicists that the fine papers used in high-quality printing, photocopying and writing are indeed radioactive.

Using standard γ -ray spectroscopic

^{40}K . The dose at the centre of the body of the person standing about 0.4 m in front of the bookcase is about $0.6 \mu\text{rad h}^{-1}$. This dose, although small, is comparable to that received by a person living or working in a brick or masonry building (in addition to the normal average background). It is interesting to note

RADIOACTIVITY LEVELS OF SELECTED PAPER SAMPLES (pCi per kg dry weight)

Sample	^{232}Th	^{226}Ra	^{40}K
Printed in US			
<i>Nature</i> (USA edn)	681 ± 40	275 ± 25	$<1,450$
<i>Science</i>	795 ± 40	300 ± 25	$<1,450$
<i>Arctic</i>	15 ± 15	20 ± 2	
<i>Time</i>	667 ± 35	182 ± 20	$<1,450$
Textbook	506 ± 25	153 ± 15	$<1,450$
Photocopy paper	266 ± 25	90 ± 20	$1,550 \pm 150$
Newsprint	45 ± 20	20 ± 20	
Printed in UK			
<i>Physics World</i>	844 ± 45	$1,310 \pm 50$	$4,550 \pm 200$
<i>J. environ. Rad.</i>	30 ± 30	50 ± 25	
Printed elsewhere			
<i>Nuclear Phys.</i>	913 ± 50	$1,195 \pm 50$	$5,150 \pm 200$
<i>Austr. J. Phys.</i>	$1,810 \pm 50$	460 ± 40	$2,800 \pm 200$

methods with a shielded large-volume HpGe detector, we have measured the γ -rays emitted by various paper samples, mostly in the form of unbound scientific journals, magazines and newsprint. In about 24 hours of counting for each sample, the main radiations from paper samples (after room background subtraction) were found to be due to ^{232}Th , ^{226}Ra and ^{40}K . Based on the observed intensity of the most intense γ -ray from each isotope, the absolute activities (pCi per kg dry weight) of selected samples are given in the table.

These data establish the presence of easily detected radionuclides in fine papers. Although we have chosen mainly scientific journals, the activities are expected to be comparable for any publications that use this type of paper. We believe that large differences in absolute activities and in thorium/radium ratios are caused by varying amounts and qualities of the fine clays that are added (a practice started about 50 years ago) to basic paper pulp to produce the durable and high-gloss surfaces used for many colour-illustrated magazines and professional journals. The thorium/radium ratio is determined by the type of fine clays available to paper manufacturers in different continents. The possibility of radioactive printer's ink is excluded by measuring blank photocopy paper.

Our data can provide an estimate of radiation exposure to a person using a library. A seven-shelf bookcase of a typical journal, for example, *Nuclear Physics*, will contain about $0.08 \mu\text{Ci}$ of ^{232}Th , $0.14 \mu\text{Ci}$ of ^{226}Ra and $0.60 \mu\text{Ci}$ of

that the dose from paper can be reduced by the use of fine papers of low radioactivity as is the case for the journals *Arctic* and the *Journal of Environmental Radiation*. Whether such adjustments to our lifestyles are justified by the dose rates involved is open to question.

B. SINGH

Tandem Accelerator Laboratory,
McMaster University,
Hamilton, Ontario,
Canada L8S 4K1

H. W. TAYLOR

Physics Department,
University of Toronto,
Mississauga,
Ontario, Canada L5L 1C6

Sexual selection and MHC genes

SIR — Potts *et al.*¹ suggest that mate preferences can account for the high genetic diversity of the major histocompatibility complex (MHC). This follows from a study of enclosed populations of wild-caught mice. In all populations there were fewer offspring homozygous for MHC than would be expected from non-random mating. The main cause of the deficiency of homozygous offspring, according to Potts *et al.*, is that when females mate extraterritorially, they prefer males that are relatively MHC-disparate. Here we suggest other interpretations that need to be excluded, and which have general implications for the study of mating preferences attributed to genetic markers.

One alternative explanation for the dearth of MHC-homozygous offspring is that MHC-heterozygous males are fitter than MHC-homozygous males (which could result from genes linked to the MHC locus). For example they may be more active or have larger home ranges. Second, females may tend to reject MHC-homozygous territorial males because they are less active. So when a female mates extraterritorially, she will have a higher encounter rate with heterozygous males, independent of any potential mate preference. Assume that mating frequency is proportional to encounter rate. If the population is in approximate Hardy-Weinberg equilibrium, then extraterritorial matings will yield more heterozygous offspring.

These possibilities can be tested. The former can be rejected if MHC-heterozygous males are not fitter in ways that affect their mating success. However, such a result is also likely to lessen one reason for mate choice, that heterozygosity at either the MHC or linked loci confers a fitness advantage. The second possibility, that females avoid mating with MHC-homozygous territorial males can be tested by comparing territorial males accepted and rejected as mates.

Doubts can also be raised over the evidence that extraterritorial matings produce an excess of MHC-heterozygous offspring. For extraterritorial matings, Potts *et al.* compare the number of heterozygous and homozygous offspring to that expected "if the female has mated solely with her territorial male". This reveals that these females have an excess of heterozygous offspring compared to that expected if they had mated with their territorial male. But this does not necessarily indicate that females mating extraterritorially prefer relatively MHC-disparate males. For that to be true, the frequency of heterozygous and homozygous offspring from extraterritorial matings should be different from the expectation for a female who engages in extraterritorial matings and who chooses her mate at random. These expected values can be derived directly from the data, which unfortunately have not yet been published.

The finding that mice can discriminate between individuals on the basis of urinary odours related to MHC alleles is suggestive that such a system could be used for mate choice in the wild. Experiments designed to demonstrate that mate choice is influenced by a particular genetic marker must consider two general, additional explanations. The first is related to our example of territory size: any behavioural factor correlated with the genetic marker that tends to make females more likely by chance to encounter males with that marker is a plausible alternative cause of nonrandom

mating. Second, females may prefer male characteristics that are indicators of the genetic marker rather than the marker itself. It is important to uncover the phenotype that females use to choose mates as well as the genetic qualities it indicates.

ANDREW POMIANKOWSKI

Department of Genetics and Biometry,
University College London,
London NW1 2HE, UK

MARK PAGEL

Department of Mathematics,
Queen Mary's College,
London E1 4NF, UK

POTTS *ET AL.* REPLY — Pomiankowski and Pagel raise some important issues which space did not permit us to address in our paper¹. First, it is important to establish that of the three kinds of mating preferences that could account for our observed deficiency of MHC homozygous progeny — preferences for heterozygotes, for rarity or for disparity — our data support only the last of these. A mating advantage for males with a rare MHC can be excluded because this hypothesis predicts an excess of rare homozygotes, whereas a deficiency is observed (in preparation).

A mating advantage for MHC-heterozygous males would, as Pomiankowski and Pagel suggest, result in a deficiency of MHC-homozygous offspring, but only if populations are in Hardy-Weinberg equilibrium; ours were not. They were founded with equal numbers of MHC heterozygotes and homozygotes to enhance tests of the overdominance hypothesis. So, even under the extreme and demonstrably false (see below) assumption that females reject all MHC-homozygous males, the expected number of MHC homozygous offspring does not differ significantly from random mating expectations. Consequently, a mating advantage for MHC-heterozygous males could not explain our observed deficiency of homozygous progeny.

In addition, the following data directly test the examples cited by Pomiankowski and Pagel. First, the mean territory size for MHC-heterozygotes (8.8 m²) does not differ from homozygotes (9.6 m²), so random encounter rates by females would accurately reflect true genotypic proportions of territorial males. In fact, random extraterritorial matings would have produced fewer MHC-heterozygous offspring due to the nonrandom settlement patterns favouring MHC-disparate pairs¹. Second, for the extraterritorially mating females in Table 3 of ref. 1, mating was random with respect to MHC homozygosity or heterozygosity of males ($\chi^2=0.05$, $P=0.82$). But these matings showed an MHC-

disassortative pattern which departed significantly from the random expectations proposed by Pomiankowski and Pagel ($\chi^2=5.02$, $P=0.025$).

Pomiankowski and Pagel are concerned that the reason for MHC disassortative mate choice is undercut if MHC heterozygotes are not fitter. In general, we agree with this concern, but it is possible that MHC heterozygotes enjoy a fitness advantage too weak to be detected with the sample size of our study². For example, a selection coefficient of $s=0.01$ operating against homozygotes would be undetectable in our system, but would be a strong evolutionary force favouring MHC-based mating preferences². Alternatively, our manipulation of genome-wide inbreeding (Table 1 of ref. 1) resulted in detectable levels of inbreeding depression. Thus, fitness consequences of MHC-based inbreeding avoidance may still be detectable in experiments similar to ours.

Finally, Pomiankowski and Pagel raise the problem of distinguishing preferences for correlated characters from characters directly dependent on MHC. This is a general problem when a single class of mates is preferred, but under disassortative mating every class of prospective mates are preferred by some class of choosers. For example, a preference for vigour could favour heterozygotes or rare males, but it could not lead to disassortative patterns because such preferences depend on the chooser's genotype.

Correlated characters arising from linked genes offer one remaining problem, which has been addressed by numerous genetically controlled experiments demonstrating that classical MHC loci are dominant in controlling individual odours³ and in influencing laboratory mating preferences^{3,4}. These experiments make use of F₂s from congenic strains, specific mutations within the MHC, cross-fostering (reviewed in ref. 3) and germ-free mice⁵. Combining these results with the considerations discussed above makes it unlikely that MHC-based mating preferences result from either correlated characters or genes other than MHC loci themselves.

WAYNE K. POTTS

C. JO. MANNING*

EDWARD K. WAKELAND

Center for Mammalian Genetics,
University of Florida,
Gainesville, Florida 32610, USA

*Also at: Department of Psychology, University of Washington, Seattle, Washington 98195, USA.

1. Potts, W. K., Manning, C. J. & Wakeland, E. K. *Nature* **352**, 619–621 (1991).
2. Potts, W. K. & Wakeland, E. K. *Trends ecol. Evol.* **5**, 181–187 (1990).
3. Boyse, E. A. *et al.* in *Psychoneuroimmunology* 2nd edn (eds Ader, R. *et al.*) 831 (Academic, New York, 1990).
4. Egid, K. & Brown, J. L. *Anim. Behav.* **8**, 548–549 (1989).
5. Yamazaki, K. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **87**, 9413–9416 (1990).

Sustainable populations?

SIR — Hoppe's provocative assertion that "Brazil has no population pressure . . ." (*Nature* **355**, 593–594; 1992) begs the question of how overpopulation can be recognized. Subjectively chosen indicators might perhaps include non-sustainable internal consumption of resources, mass starvation or unemployment.

Ecologists, economists and politicians might have different opinions as to whether an area is, or will become, overpopulated. I trust that most geologists, and Hoppe, could recognize population pressure long before the density of people is sufficient to inhibit their mining projects!

The crucial but controversial issue of the sustainable population of individual countries, and of the globe, demands international and interdisciplinary agreement; the forthcoming environmental summit in Brazil should be a suitable venue for discussion.

CLIVE HAMBLER

Department of Zoology,
University of Oxford,
Oxford OX1 3PS, UK

SIR — Hoppe¹ is wrong in his view that Amazonian rain forests cannot be exploited sustainably. Tropical rain forests, just like others, are dynamic ecosystems². Trees are continuously dying individually or in groups to create canopy gaps that are the first stage of a growth cycle. Gaps are colonized by seedlings that grow up eventually to form a new mature forest. This is well known to foresters who have developed silvicultural systems that work within the inherent biological limits and allow the harvesting of trees as they mature.

Such systems are well established in the Asian rain forests³ and have recently been devised for the neotropics^{4,5}. Economic exploitation is sustainable so long as it is carefully conducted to minimize erosion, soil compaction and damage to juvenile trees that will form the next growth cycle. Careless or excessively heavy exploitation delays recovery but does not destroy forest. What does destroy forest is its conversion to either crops or pastures. For timber extraction roads are built and often destruction occurs because these give access to

1. Hoppe, E. *Nature* **355**, 593–594 (1992).
2. Whitmore, T. C. *Introduction to Tropical Rain Forests* (Clarendon, Oxford, 1990).
3. Wyatt-Smith, J. *Malayan Forest Records* **23**, 1963.
4. de Graaf, N. R. A. *Silvicultural System for Natural Regeneration of Tropical Rain Forest in Suriname* (Agricultural University, Wageningen, 1986).
5. Silva, J. N. M. *The Behaviour of the Tropical Rain Forest of the Brazilian Amazon after Logging* thesis (University of Oxford, 1989).

farmers, but this is a social and political problem not connected to forest sustainability.

T. C. WHITMORE

Geography Department,
Downing Place,
University of Cambridge,
Cambridge CB2 3EN, UK

Concentrating mammalian urine

SIR — Alexander in *News and Views*¹ asserts that "The Australian hopping mouse (*Notomys*) conserves water . . . by producing more concentrated urine than any other mammal. . . To do this, it needs very long kidney tubules. . .". Actually, small rodents, including *Notomys*, concentrate urine with very short kidney tubules. In *Notomys*, the longest loops of Henle, the tubules responsible for concentrating urine, are only 5.2 mm long, and *Notomys* achieves a urine concentration of 9,370 mosmol. Horses, with loops 7 times as long, achieve maximum concentrations of only 1,900 mosmol. For all mammals, there is no trend for increased concentrating ability with increased loop of Henle length². Birds decrease urine concentrating ability with increased loop length³.

Small mammals with short loops might produce concentrated urine because of the relationship between mass-specific metabolic rate and body size^{4,5}. Small mammals (and their kidneys) have more intense metabolic rates than large mammals. The mitochondrial density in the loops of Henle in small mammals (such as mice) is about twice the density of large mammals (horses), and renal mitochondria in small mammals are more densely packed with cristae than in large mammals⁶. Tubule cells of small mammals also have higher densities of basolateral membrane infoldings (insertion sites for the ion pumps that contribute to concentrating urine). Thus, small mammals may produce high urine concentrations with short loops because, compared to larger mammals, the kidneys of small mammals may have a greater capacity for the active transport underlying the concentrating mechanism.

Alexander asks why other mammals do not match the concentrating ability of desert rodents, yet few desert rodents exceed the concentrating ability of the house mouse (*Mus musculus*), not a desert specialist⁷. Why large mammals do not have greater renal mitochondrial densities may be a matter of rules relating body size to metabolic rate. Given the scant evidence relating tubule length to concentrating ability, it is difficult to understand why any mammal has any

particular loop length. In sum, long loops of Henle exist more in the minds of zoologists than in the kidneys of desert rodents.

LEWIS GREENWALD

Department of Zoology,
Ohio State University,
Columbus,
Ohio 43210-1293, USA

1. Alexander, R. McN. *Nature* **353**, 696 (1991).
2. Beuchat, C. A. *Am. J. Physiol.* **258**, R298-R308 (1990).
3. Goldstein, D. L. & Braun, E. J. *Am. J. Physiol.* **256**, R501-R509 (1987).
4. Greenwald, L. *Physiol. Zool.* **62**, 1005-1014 (1989).
5. Greenwald, L. & Stetson, D. *News Physiol. Sci.* **3**, 46-49 (1988).
6. Abrahams, S., Greenwald, L. & Stetson, D. L. *Am. J. Physiol.* **261**, R719-R726 (1991).
7. Haines, H., Ciskowski, C. & Harms, V. *Physiol. Zool.* **46**, 110-128 (1973).

Plants at the K/T boundary

SIR — I cannot accept the scheme presented by Jack A. Wolfe¹ for a 'June impact' at the Cretaceous/Tertiary boundary. Not only has Wolfe extrapolated an elaborate sequence of events of global consequence from a single locality, but some of the critical evidence he cites is in my view invalid.

Wolfe reports two kinds of fossil pollen from the Teapot Dome locality that he interprets respectively as "Nuphar-like pollen" (Fig. 3e) and "Nelumbo-like pollen" (Fig. 3f). On the basis of my inspection of Wolfe's slides, which he made available to me before publication of his letter, I believe that both kinds of pollen were misidentified. Specimens of alleged pond lily (Nymphaeaceae) pollen (Wolfe's Fig. 3e) are not "Nuphar-like" because they lack characteristic monosulcate apertures. They are *Pandaniidites typicus* (Norton) Sweet (synonym: *P. radicus* Leffingwell), a species having a monoporate aperture; this species has probable affinity with the Pandanaceae². It is common in uppermost Cretaceous (Maastrichtian) as well as lower Tertiary (Palaeocene) rocks of the region^{3,4}, hence its stratigraphic range exceeds that of the solely Palaeocene leaf species *Paranymphaea crassifolia*. Wolfe's argument that the co-occurrence of the leaves and pollen in his samples indicates derivation from the same plant is negated by the significant discordance in their total stratigraphical ranges.

Specimens of alleged lotus (Nelumbaceae) pollen such as the tetrad illustrated by Wolfe (Fig. 3f) are not "Nelumbo-like" because they are tetrads (extant *Nelumbo* produces pollen as individual grains or monads of considerably larger size than the fossils) and because the fossils lack the tricolpate apertures and complexly structured walls (exines) characteristic of *Nelumbo* pollen. These fossils are *Inaperturotetradites*

scabratus Tschudy, a species of uncertain botanical affinity but lacking any morphologic resemblance to pollen of *Nelumbo*. Wolfe's statement that "in extant *Nelumbo*, tetrads occur in young (immature) pollen" is irrelevant, because pollen of all species of seed plants originates as tetrads (as a consequence of meiosis during pollen genesis). Wolfe seems to have misinterpreted the caption to a drawing by Erdtman⁵ illustrating the relative positions of colpate apertures on adjacent daughter cells of a tetrad during development of the pollen. Erdtman's drawing clearly illustrates the well-developed tricolpate apertures and complex exine of *Nelumbo* pollen — features lacking in the inaperturate, thin-walled fossils in the Teapot Dome samples.

More significant than these misidentifications of the fossil pollen is the observation that fossil tetrads identical to those from Teapot Dome, which Wolfe regards as immature stages from plants suddenly frozen at the Cretaceous/Tertiary boundary, are widely known from Upper Cretaceous (Campanian and Maastrichtian) rocks in Wyoming and Montana⁶. Unless countless sudden freezing events throughout Late Cretaceous time are invoked to explain other occurrences of *I. scabratus*, the Teapot Dome specimens cannot be said to represent pollen frozen during the terminal Cretaceous event. Without immature "Nelumbo-like" pollen to indicate unopened flowers of fossil *Nelumbites*, the hypothesis that an impact happened in June at the end of Cretaceous time has no support.

Further, my counts of fossil pollen on Wolfe's slides revealed that the tetrads in question are most abundant in bed 2 and that the misidentified specimens said to represent pond lilies are most abundant in bed 4. The law of superposition dictates that the tetrads were produced before the alleged lily pollen, the reverse of the sequence listed by Wolfe in his Table 1. Therefore, although Wolfe's analysis suggests the potential that exists for detailed analyses of the stratigraphic record at the Cretaceous/Tertiary boundary, his interpretation of the events that occurred are not supported by the available data.

DOUGLAS J. NICHOLS

US Geological Survey,
Mail Stop 919, Box 25046,
Denver, Colorado 80225, USA

SIR — Despite calling attention to the potential of microstratigraphic analysis to provide insights into geological events of extremely short duration, the letter by J. A. Wolfe¹ contains numerous palaeobotanical and stratigraphic statements to which we object.

Although fresh leaves of *Nuphar* that we froze for more than 3 weeks dupli-

cated the pattern of fine-scale cuticular folding illustrated by Wolfe, identical patterns also appeared on *Nelumbo* and *Nuphar* leaves affixed to herbarium sheets, as well as in the epidermis of both previously frozen and unfrozen leaves allowed to dry on glass slides. The folding varied in its extent and was most commonly found near the margin, larger veins and folds, but was not seen on the cuticle of decaying *Nuphar* leaves taken directly from a water bath. After death, the mesophyll of both *Nuphar* and *Nelumbo* could be seen to undergo rapid breakdown into a gelatinous mass. From these observations we infer that such folding is the result of boundary shear stress generated by the slippage of the degrading mesophyll over the more integral epidermis, whose movement is constrained in some way, as by the leaf margin, veins or by adhesion to an interface such as that with the enclosing sediment. Thus we believe that Wolfe's implication that freezing is a unique cause of his cuticular folding is not correct.

We also take issue with some of Wolfe's palaeobotanical identifications and biostratigraphical assertions: for example, leaves of the form-genus *Nelumbites* lack the hexagonal areolation possessed by all modern species of the genus *Nelumbo*; the alleged seed shown in Fig. 3k cannot be positively identified; and the elements in Wolfe's Fig. 3h cannot be reliably recognized either as "growing tips" or "Nelumbonaceae" as labelled. Further, leaves of the extinct form *Paranymphaea crassifolia* differ significantly from those of *Nuphar* in secondary vein spacing and areolation. It thus seems rash to ally *Paranymphaea* with *Nuphar* and to attribute the climatic tolerance and even time of blooming and seed-set of *Nuphar* to the fossil. We also question Wolfe's ability to determine mean *daily* temperatures using unstated methods, especially in view of the significant objections that have been raised to his leaf physiognomic methods for determining mean *annual* temperatures⁷.

Far from being anomalous, ponded-water sediments are common⁸ and the assemblage of plants that Wolfe describes is a normal part of the basal Palaeocene sequence for tens of metres or more above the Cretaceous/Tertiary boundary in this region^{9,10}. Despite an extensive fossil record, *Paranymphaea crassifolia* has not yet been reliably identified anywhere in Cretaceous sediments. The assertion that the rhizomes of *Paranymphaea* and *Nelumbites* were lifted out of their growth position in the latest Cretaceous mudstone beneath Wolfe's bed 1 and redeposited in bed 5 of the sequence (Palaeocene) by freezing of the higher parts of the plants into an ice layer that was later buoyed by flood

waters appears highly improbable in view of the lack of disruption of the Cretaceous mudstone or any evidence of remaining leaves or rhizomes in it. Additionally, our observations show that the leaves, petioles and rhizome tips of the modern analogues lose all fundamental strength upon being frozen or shortly after death, and that the leaves sink rapidly after death and are readily disaggregated after as little as 1–3 weeks at 20 °C in the laboratory. Thus, it is highly unlikely that the leaves would have continued floating long after thawing or could have had the coherency to collect sediment on their upper surfaces.

LEO J. HICKEY

LUCINDA J. McWEENEY

Department of Geology and Geophysics
and Department of Anthropology,
PO Box 6666,
170 Whitney Avenue,
New Haven,
Connecticut 06517, USA

WOLFE REPLIES — I proposed¹ a scheme for Cretaceous/Tertiary boundary events based on the Teapot Dome section, which represents an *in situ* lily pond and is the only known boundary section that contains determinable plant megafossils in the bolide fallout layers. All megafossil taxa, which represent pond lily (*Paranymphaea*) and lotus (*Nelumbites*), were discussed and illustrated; these megafossils were interpreted as remains of Cretaceous plants that suffered mass-kill. I reported a series of laboratory experiments on modern lotus (*Nelumbo nucifera*) leaves; these experiments, documented by photomicrographs, indicate that structural deformation like that in the fossil cuticle was produced by freezing but not by other environmental disturbances, including dessication. Further, the deformed cuticles are restricted to the fallout layers, which contain no evidence of dessication. Only two types of pollen occur in antherial masses and hence probably came from plants proximal to the depositional site; these pollen have characters that I interpret to ally them to extant relatives of the megafossil taxa.

Is *Paranymphaea crassifolia* in fact absent in the Cretaceous? If true, the *Paranymphaea* leaves I found 0–5 cm above the impact layer could not represent Cretaceous plants, and *P. crassifolia* must have originated *de novo* at the boundary, a most remarkable and improbable event.

Pollen morphology can be difficult to interpret. Nichols and I interpret the morphology of the spinose Teapot Dome pollen differently, and he asserts that this pollen cannot be nymphaeaceous. This pollen consistently has a single 'fold' and, despite my requests, Nichols was unable to find a 'pore' that was not

obscured by a spine. I interpret this pollen as monosulcate and nonporate. Nichols also seemingly does not appreciate that different plant organs can evolve at different rates; the nelumbonaceous pollen, which is distinct from the Cretaceous taxon he mentions, does not have all specialized characters of extant *Nelumbo*, unlike the associated megafossil organs. He denies the significance of the occurrence of nelumbonaceous pollen as tetrads; if this is not significant, why did the typically terse Erdtman⁵ mention this? Nichols' palynological arguments result in the improbable conclusion that in this lily pond neither *Paranymphaea* nor *Nelumbites* was represented by pollen.

If Hickey and McWeeney are attempting to show that my statements on cuticle are false, why did they experiment on *Nuphar* and not *Nelumbo*? *Nelumbites* constitutes about 98% of the megafossils, strongly suggesting that this plant was the source for the structurally deformed fossil cuticle. To observe the minor folds, I had to prepare the leaves chemically, whereas Hickey and McWeeney state that such folds "... appeared on *Nelumbo* and *Nuphar* leaves affixed to herbarium sheets ...". Hickey and McWeeney's folds varied in distribution, whereas my freezing folds are ubiquitous in modern and fossil cuticles. I question whether Hickey and McWeeney are reporting the same kind of structural deformation. I also question the observations on the supposed lack of strength of aquatic organs following freezing: in my experiments, thawed *Nelumbo* leaves underwent some degradation but remained intact and floating for at least 4 months.

Determination of events at the Cretaceous/Tertiary boundary is important, and I thus allowed Nichols to examine my preparations before publication. Similarly, had they asked, Hickey and McWeeney could have examined the megafossils. Why do they now contradict my megafossil determinations without having examined the specimens?

JACK A. WOLFE

US Geological Survey,
Mail Stop 919, Box 25046,
Denver,
Colorado 80225-0046, USA

1. Wolfe, J. A. *Nature* **352**, 420–423 (1991).
2. Muller, J. *Bot. Rev.* **47**, 1–142 (1981).
3. Leffingwell, H. A. *Geol. Soc. Am. spec. Pap.* **127**, 1–64 (1970).
4. Jerzykiewicz, T. & Sweet, A. R. *Can. J. Earth Sci.* **23**, 1356–1374 (1986).
5. Erdtman, G. *Pollen Morphology and Plant Taxonomy* (Chronica Botanica, Waltham, Massachusetts, 1952).
6. Tschudy, B. D. *US Geol. Surv. Prof. Pap.* **770** (1973).
7. Dolph, G. *Proc. Indiana Acad. Sci.* **99**, 1–10 (1990).
8. Fastovsky, D. E. & McWeeney, K. *Geol. Soc. Am. Bull.* **99**, 66–76 (1987).
9. Hickey, L. J. *Univ. Mich. Pap. on Paleont.* **24**, 33–49 (1980).
10. Johnson, K. R. & Hickey, L. J. *Geol. Soc. Am. spec. Pap.* **247**, 433–443 (1991).

Britain's Chernobyl?

John Simpson

Windscale 1957: Anatomy of a Nuclear Accident. By Lorna Arnold. *Macmillan/St Martin's Press*: 1992. Pp. 235. £40, \$55.

"A MONUMENT to our ignorance" was how Sir Christopher Hinton described the United Kingdom's first nuclear reactors dedicated to military purposes. But Lorna Arnold's pithy, authoritative and incisive account of the events surrounding the fire in the No.1 pile at Windscale (now Sellafield) in northwest England on 10–11 October 1957 indicates that this ignorance embraced both the behaviour of graphite when irradiated and heated

error' was officially given as its cause, the failure to investigate the effects of irradiation and heat on graphite, and the reasons for the absence of a 'safety culture' in the British nuclear programme of the 1950s. Her key conclusion is that the cause of the fire was not the bursting of a uranium or a lithium-magnesium cartridge because of overheating, as suggested in the official reports around the time, but narrowing

inherited from wartime, that top priority had to be given to meeting military production targets. Taking risks in such an atmosphere was inevitable. The situation was compounded by several other factors. One was a shortage of senior staff. Although the creation of the United Kingdom Atomic Energy Authority (UKAEA) in 1954 was supposed to ameliorate the shortages of senior engineers and managers by enabling competitive salaries to be paid, shortages still continued as salaries remained linked to those of the civil service.

A second factor was the lack of clear lines of communication and responsibility in the UKAEA, allied to a parochial tradition of individual plants solving their own problems. This made it difficult to create a safety culture within the

BNFL

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

and the need for nuclear programmes to be embedded in an effective safety culture, rather than in a 'can-do' one.

The two Windscale piles provided plutonium and polonium-210 for Britain's first atomic explosive devices, and tritium for its later thermonuclear ones. Until late 1956 they were the only military-production reactors in the United Kingdom, and thus any suspension of their operations before 1957 would have drastically curtailed the UK thermonuclear weapon development programme. The pile design involved the blowing of cooling air over natural uranium fuel in magnesium cartridges emplaced within a graphite structure. To prevent the cartridges from bursting, the graphite had to be 'annealed' by heating. Otherwise irradiation would result in the build-up of strain in the graphite lattice, with the possibility of a sudden release of Wigner energy leading to an unpredictable increase in temperature. It was while the annealing process was taking place, with the fuel providing a low level of nuclear heating and the cooling blowers switched off, that the fire occurred.

Arnold provides new information and analysis on several aspects of the Windscale fire: the immediate causes of the fire, the political reasons why 'human

through irradiation of the gap between the temperature at which the graphite would smoulder and catch fire and that reached in the annealing process. Eventually a point was reached where a fire was inevitable whatever the operators did. This leads to two crucial questions: why was human error implied as the cause in the White Paper (policy document) published in the aftermath of the accident, and why was there so much contemporary ignorance about the effects of irradiation on graphite?

Sir William Penny's initial report contained no attribution of blame. It was Prime Minister Harold Macmillan's White Paper of November 1957 that gave the cause as "partly due . . . to faults of judgement by the operating staff". The explanation lies in his own words: "The publication of the report, as it stands, might put in jeopardy our chance of getting Congress to agree to the President's proposal". This proposal was that nuclear energy collaboration for military purposes between the United States and the United Kingdom should be restarted after 13 years.

This economy with the truth masked a deeper cause of the fire, which it is difficult fully to appreciate and comprehend in the world of 1992: the ethos,

overall organization, or to give it a high enough priority. A third factor was that research and development were focused on many new projects, and possible problems with the old, obsolescent Windscale piles were largely ignored. This explains why so little work was done on the effects of irradiation on the graphite in the piles. Moreover, Arnold makes clear the high technical risks involved in the 1955 decision to build civil nuclear power reactors on the basis of the design at the Calder Hall plant. (In April 1958, it was still thought possible that the graphite in these reactors would need annealing and that they might have a working life as short as seven years. Yet despite the uncertainties, the massive programme continued.)

Some will no doubt use Arnold's findings to construct a conspiracy theory that senior management and politicians were aware of the risks of running the piles, but continued to do so because of the priority given to developing thermonuclear weapons. The evidence produced by Arnold suggests that this was not so.

The Windscale fire brought policy-makers face to face with the fact that they were demanding too much of the UKAEA and providing it with too few resources. Fortunately for them, the fire

happened after the United Kingdom's thermonuclear explosive capabilities had been demonstrated, and coincided with the decision of the president of the United States to attempt to restart military nuclear collaboration. It also led to the exchange of a can-do culture within the programme for a strengthened safety one. In view of the Chernobyl experience, Britain may count itself fortunate that this happened before its civil nuclear power reactors were commissioned, and with such minimal environmental and medical consequences. □

John Simpson is at the Mountbatten Centre for International Studies, Department of Politics, University of Southampton, Southampton SO9 5NH, UK.

Filling in the gaps

David N. Schramm

The Hidden Universe. By Roger J. Tayler. Ellis Horwood: 1991. Pp. 213. £30 (hbk), £15.99 (pbk).

COSMOLOGY has become one of the most active and exciting topics in modern physical science. Never before have so many experiments and observations had possible cosmological implications. Although modern cosmology probably started in 1929 with Edwin Hubble's observations of an expanding Universe, the next important observation of immense cosmological impact was not until the discovery by Arno Penzias and Robert Wilson in 1965 of the microwave background radiation. In the past decade the pace of discovery has quickened enormously, with accelerator experiments and telescopes on the ground, as well as satellites in space, giving us information that has cosmological relevance. This has enabled cosmological studies to move from the more philosophical side of intellectual endeavour to become a genuine physical science where hypotheses are checked by making predictions that lead to actual measurements.

Although my personal interest no doubt biases my view, I feel it is safe to say that, along with the microwave background, the light elements have had the most lasting recent effects. The cosmological nucleosynthesis predictions for the abundances of the light elements, ^1H , ^2H , ^3He , ^4He and ^7Li , span a range of almost 10 orders of magnitude and each is now observed, in primitive material, to be at its predicted value. Furthermore, it was shown that the amount of ^4He is related to the number of types of fundamental parti-

cles. In particular, we now know that there could be no more than three families of neutrinos without exceeding the observed primordial abundance of helium. This number of neutrino families was directly observed last year at the LEP collider at the European Laboratory for Particle Physics (CERN) in Geneva, Switzerland. So cosmology is now making predictions that can be checked with particle accelerators as well as telescopes.

Roger Tayler was one of the pioneers in cosmological nucleosynthesis calculations. Even before the microwave background was first measured, Tayler's work with Fred Hoyle showed that primordial nucleosynthesis would convert an appreciable fraction of the mass of the Universe into ^4He . Showing foresight, they even noted at the time a connection between the helium yield and neutrinos, although they did not explicitly calculate anything about this relationship. Clearly, Tayler has been one of the key players in the recent developments in cosmology.

Although there have been many books written over the past few years that have tried to convey the current level of excitement in the field to the nonscientifically educated public, little has been written for the more mathematically literate nonspecialist. *The Hidden Universe* nicely fills this need. Like several other recent books, it is an excellent and very readable presentation of the main topics in modern cosmology. But it differs from the others in that it is not afraid to use algebra. This, unfortunately, distances the book from the general public, but it makes it much more useful for scientists in other fields who wish to gain an insight into the excitement of cosmology today. Furthermore, the book is not meant as an advanced text or monograph for graduate students. Instead, it is written at a level that undergraduates or even advanced secondary school students can understand as long as they are familiar with elementary algebra.

Tayler presents the basic development of the current cosmological picture, leading his readers naturally into the central problems in current cosmology — namely, what constitutes the so-called dark matter in the Universe and how does large-scale structure, such as galaxies and 'great walls', arise from the apparently 'smooth' early Universe.

The need for some sort of exotic dark matter is a straightforward consequence of the light-element story. The light-element abundances all give the remarkable fit mentioned above, but only if the density of ordinary neutron- and proton-based matter is relatively low. But other arguments imply that the overall density of the Universe is much higher, hence

the need for some new exotic stuff. Tayler's arguments are not watered down, but are fully presented with the appropriate algebraic equations so that they can be followed.

A disappointment with the book is its lack of references. Although there is a good list of annotated suggestions for further reading in the back pages, there is no mention of where to go for further details on specific points. This drawback could limit the book's usefulness as a text in university courses. Tayler does in fact admit in the introduction that there will be very few citations of scientific contributions, but this sparseness is very noticeable when figures, taken from other published sources, are not fully cited. By contrast, the appendices, on such topics as degenerate matter, the virial theorem and thermodynamic equilibrium, present short, simple derivations of useful physics that are ideal for undergraduates or for scientists from other fields who need to refresh their memories. In short, *The Hidden Universe* provides a very readable and much needed introduction to modern cosmology for any algebraically literate reader. □

David N. Schramm is in the Department of Physical Sciences, University of Chicago, Chicago, Illinois 60637, USA.

Chemical controversies

W. P. Griffith

Oxygen Chemistry. By Donald T. Sawyer. Oxford University Press: 1992. Pp. 223. £30, \$35.

THIS interesting and stimulating book represents very much an individual view. Even the first sentence in the preface could be taken by some chemists as controversial: "The book results from a conviction that oxygen is the most important element in the realm of chemistry". As R. J. P. Williams says in a perceptive introduction, "The author probably knows that he is being provocative [but he] is writing from deep knowledge and experience".

Sawyer rightly emphasizes the central role of oxygen in chemistry and provides a good overview of the principal oxygen-containing species (dioxygen, ozone, hydrogen peroxide, alkyl hydroperoxides, peracids; the hydroxyl, perhydroxyl, hydroperoxo, ozonide, oxo and superoxo anions, oxyanions; and also oxygen-containing radicals). The focus, appropriately, is on the reactivity of these species and on the kinetic and thermodynamic data that have been collected for them. In a rather curious and

idiosyncratic third chapter on bonding in some of these species, Sawyer rejects some traditionally held views on ionicity in oxygen-containing species. This treatment will provoke a hostile reaction in some readers. Still, Sawyer has never been averse to chemical controversy.

This is very much a book for the specialist rather than the general reader — one looks in vain for any mention of the causes of ozone depletion or accretion in the upper or lower atmospheres, for example. Tantalizing glimpses are provided of the involvement of oxygen and its compounds with biological systems, but this is not a theme developed to any depth (although, to be fair, Sawyer clearly recognizes the crucial importance of oxygen in biological chemistry and urges the reader to follow up these topics elsewhere). Catalysis, a vital subject in any treatment of oxygen chemistry, is also given a rather partial treatment. Manganese-, iron- and cobalt-induced catalysis is well repre-

sented, but there is no mention at all of the academically fascinating and industrially important activation of hydrogen peroxide by the early transition metals, or of peroxide activation by the platinum-group metals. Nevertheless, the book contains a wealth of information and is recommended to all chemists interested in oxygen chemistry and requiring a little intellectual stimulation.

Although the book is well bound and, by modern standards, reasonably priced, the typeface is not aesthetically pleasing and shows a curious, intermittent irregularity in line spacing. A publishing house as distinguished as OUP surely could do better than this. Still, as befits a book on what is undoubtedly the most environmentally friendly of all elements, it is good to know that it has been printed on acid-free paper. □

W. P. Griffith is in the Department of Chemistry, Imperial College of Science, Technology and Medicine, London SW7 2AY, UK.

High-flyer

R. V. Jones

Boffin: A Personal Story of the Early Days of Radar, Radio-astronomy and Quantum Optics. By R. Hanbury Brown. Adam Hilger/AIP: 1991. Pp. 184. \$35, £17.50.

IN 1935, Henry Tizard, who was chairman of the Committee for the Scientific Survey of Air Defence and also Rector of Imperial College, recruited one of the college's research students in electrical engineering to join the nascent radar project under Robert Watson Watt at Bawdsey. The student was Robert Hanbury Brown who, although not yet 20 years old, had both a first class honours degree in electrical engineering and a pilot's licence gained by flying with the London University Air Squadron.

Coming from a family with a military background and roots in Sussex going back some five hundred years, and with the best attributes of an English gentleman overlaid by high spirits and inexhaustible energy, Hanbury, as we came to call him, rapidly stood out among the devoted team who were working frantically in spartan conditions to develop radar in time for the impending war. Most of the effort went into the development of ground-based radar (CH) to support Fighter Command by day; but it was becoming clear that this could not be precise enough to direct a fighter to within visual range of a bomber in darkness. Something extra was needed, and the most promising approach was to build an item of radar equipment small enough to go into a night-fighter to

enable it finally to close with the bomber.

In 1937 Hanbury joined a small team with E. G. Bowen and A. G. Touch to develop such an airborne radar. Within three years the team brought the device to the operational stage as AI Mark IV, whose first success came in November 1940 when a night-fighter so equipped shot down a German bomber. In the meantime, a maritime version enabled surfaced U-boats to be detected at night from the air, with vital consequences in the Battle of the Atlantic. (The later history of airborne radar is recounted more fully by Bernard Lovell in *Echoes of War*, which I reviewed in these pages in January.)

All this work was done at 1.5-metre wavelength, and Hanbury's special contribution came partly from his training as an engineer, where he was critical of the 'bread board' approach when a well-engineered alternative was available, and partly from his flying experience, where he more than anyone else could uncover the difficulties of bringing his fighter onto the tail of a bomber. Flying exacted a personal toll: the many hours at high altitude damaged his ears, and an oxygen-supply failure that rendered him unconscious and put him into hospital terminated his experimental flying in 1941.

By then, the successes at Bawdsey had enabled Britain to stave off defeat in 1940, long enough for many more scientists to join after the outbreak of war, and for the cavity magnetron and centimetric radar to be developed.

At the end of the war Hanbury, who in the meantime had been described as the original 'boffin' by Watson-Watt, joined him and B. V. Bowden in a



R. Hanbury Brown with C. Hazard examining the output of the 218-foot 'paraboloid'.

consultancy partnership that, among other things, advised film studios on such matters as how to keep the water clear in the tank created for the film *Miranda* in which the mermaid (Glynis Johns) had to swim, and the most appropriate material from which to construct God's throne for a film depicting heaven — Perspex or Lucite was the consultancy's recommendation.

Diverting as such problems were, they were not as attractive as those in the budding branch of radioastronomy, and Hanbury accordingly joined Bernard Lovell in Manchester in 1949, and immediately started work on 'radio stars' with C. Hazard. In the course of devising ways by which the stars' angular sizes could be estimated, Hanbury thought of a new type of interferometer that did not depend on correlating phases and amplitudes at two separate points, as Michelson had done at Mount Wilson with his stellar interferometer in 1920, but on correlating intensities alone.

Hanbury arrived at the proposal by considering the problem as one of radio-wave propagation, and this suggested that the technique could be applied to light waves as well. Although this aroused objections on the grounds that it would not apply to photons, he succeeded with R. Q. Twiss in showing both in theory and by experiment that there was correlation between individual photons. This was a basic step in quantum optics; moreover he demonstrated in Manchester that the intensity interferometer could be made to work on stars, and in 1962 he moved to Australia to set up a large-scale interferometer, with which he measured the angular diameters of stars well beyond what Michelson had been able to do. The success of the interferometer (which had

to be erected and operated at a remote site 360 miles from Sydney, and whose 22-foot mirrors were carried on a circular track 618 feet in diameter) was largely due to the personal effort that he put into its construction — the same factor that had led to the earlier success in airborne radar.

Hanbury once said that Watson-Watt never said in one word what could be said in a thousand. The inverse would be true of Hanbury himself, for in this slim volume is distilled much scientific wisdom and practical and operational experience. Together with E. G. Bowen's *Radar Days* (Adam Hilger, 1987) it provides a first-hand account of the early days of radar, and of the small but brilliant team who contributed so much to Britain's survival in 1940, and who carried their successes forward from

war into astronomy.

Hanbury himself strikes a resonance with Aeschylus who, despite his successes as a playwright, wished to be remembered simply for the fact that he had fought at the Battle of Marathon. 'Marathon' for Hanbury was 1940, and despite his scintillating contributions both to astronomy and to basic physics, it is of his early work on airborne radar that he writes, "Nothing which I have done since then has been so exciting, so absorbing or so worthwhile" — and, we may add, so worthy of high record in the annals of 'The Few'. □

R. V. Jones, from 1939 to 1945 head of scientific intelligence on Britain's Air Staff and scientific adviser to the British Secret Intelligence Service (MI6), is at 8 Queen's Terrace, Aberdeen AB1 1XL, UK.

Protection postponed

Len Goodwin

The Malaria Capers: More Tales of Parasites and People, Research and Reality. By Robert S. Desowitz. Norton: 1991. Pp. 288. \$21.95, £14.95.

Malaria: Waiting for the Vaccine. Edited by Geoffrey Targett. Wiley: 1991. Pp. 224. £39.50, \$84.05.

ROBERT Desowitz was born in New York, studied medical parasitology at the London School of Hygiene and Tropical Medicine and has since then spent 40 years working in many tropical countries. He is at present a professor in the University of Hawaii and lives in Honolulu; he has written a good textbook of parasitology, as well as books in which he has shared his thoughts with the general public. This is his third popular book of racy, readable stories about diseases and medicine in developing countries, backed up with yarns about his colleagues and about his personal experiences.

He traces the long struggle to discover the cause and mode of transmission of malaria, from Hippocrates to Laveran, Manson and Ross, and the fierce arguments about the best method of control — giving antimalarial drugs or eliminating the mosquito vector.

Both approaches have now been tried and both have failed. The cheap and once-effective drug chloroquine is now useless in many areas because the malaria parasites there have become resistant to it, mainly through careless underdosing. The World Health Organization's multi-billion-dollar "Global Malaria Eradication Programme", in which the walls of houses were sprayed with DDT, be-

gan in 1955 and was finally abandoned in 1976; mosquitoes had become resistant or had changed their habits, and more species of mosquito, some with a frightening reproductive capacity, were found to be transmitting the disease.

In the years when DDT was working, sandflies, the vectors of another protozoal infection, kala-azar (leishmaniasis), were also killed and transmission of the disease was greatly reduced. The first part of the book deals with this disease, epidemics of which repeatedly swept through India at the turn of the last century.

Now both malaria and kala-azar are back; malaria infects 200–300 million people and causes 1–2 million deaths every year. Desowitz tries to find reasons for the continued failures of expensive international efforts at control. He blames arrogant, ignorant or incompetent leadership that disregards clear messages from scientists working in the field. Massive programmes tend to acquire massive momentum and are difficult to slow, stop or change in direction. Epic proportions are reached in the story of the malaria vaccine. For 25 years the American AID programme spent millions of dollars in grants to researchers for the preparation of a malaria vaccine (difficult, even with genetic engineering, because there are several species and many strains of parasite, all with antigenically different stages in their mammalian and mosquito hosts). Funding of projects continued in spite of clear warnings from other scientists. Four scientists were indicted for stealing large amounts of the grant money, and there is still no vaccine. Desowitz believes that there is "too large a gap between the high intellectual world of biomedical research and the needs of the sick and those who attend them."

But what now? *Malaria: Waiting for the Vaccine* is the report of a public

health forum held last year at the London School of Hygiene and Tropical Medicine, with contributions from the British Overseas Development Administration, the World Health Organization and the World Bank. New problems in the management of patients, drug distribution and use, drug resistance, environmental management and the economics and organization of various control measures are surveyed by distinguished contributors; each report is followed by a summary of workshop discussions and recommendations. David Bradley provides an excellent historical background, leading from attempts at control and eradication to resurgence, chaos and hopes for the future. "Chaos" certainly applies to the present distribution and use of antimalarial drugs; in some of the endemic countries, the malaria services, through "bureaucracy preoccupied with self-preservation", have become counterproductive, and nearly half of all antimalarial drugs are bought casually from shopkeepers and street traders — not a pretty picture.

No control method is free from problems. An effective vaccine could transform the situation, but there is much scope for the re-use and improvement of methods that have been tested in the past. A report from China shows that bed-nets impregnated with pyrethroid insecticide greatly reduce the incidence of malaria.

The World Health Organization means business and its representatives outline four principle objectives for action: timely, adequate diagnosis and treatment; vector control where practicable and cost-effective; early warning of epidemics; and regular monitoring of conditions in the field. A global conference, at ministerial level, will be held in Amsterdam in October 1992 to focus the attention of endemic countries and donor agencies on the problem.

And the vaccine? Research, of course, goes on, but fundamental questions about humoral and cellular responses to malaria antigens are as yet unanswered. We still do not fully understand the mechanisms by which immunity leads to protection against any stage of the malaria parasite. In a postscript, Lou Miller warns that although children continue to die from malaria in Africa and India — and it is difficult to accept the slow and unpredictable course of basic research — the field needs radically new control methods. Chloroquine "took hundreds of years to develop but none today would argue that the effort wasn't worth it". □

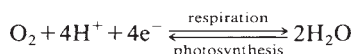
Len Goodwin (formerly with the Wellcome Trust) is at Shepperlands Farm, Park Lane, Finchampstead, Wokingham, Berkshire RG11 4QF, UK.

Oxygen activation and the conservation of energy in cell respiration

Gerald T. Babcock & Mårten Wikström

Many of the membrane-associated oxidases that catalyse respiratory reduction of O_2 to water simultaneously couple this exergonic reaction to the translocation of protons across the inner mitochondrial membrane, or the cell membrane in prokaryotes, a process by which metabolic energy is conserved for subsequent synthesis of ATP. The molecular mechanism of O_2 reduction and its linkage to H^+ translocation are now emerging. The bimetallic haem iron-copper reaction centre in this family of enzymes is the critical structure for catalysis of both these processes.

DIOXYGEN has been harnessed effectively as the terminal oxidant in aerobic respiration in a reaction that forms half the water-oxygen cycle in nature. Interconversion between these two compounds presents a fascinating example of biological control involving four-electron/four-proton oxidoreduction chemistry essential to both photosynthesis and respiration.



Cytochrome and quinol oxidases provide a sink for sustained electron transfer from foodstuffs through respiratory chains by catalysing efficient O_2 reduction. The exergonic electron transfers are coupled to proton translocation across the membrane, whereby energy is primarily stored as an electrochemical proton gradient, before its use in the synthesis of ATP¹. In fact, the principal family of oxidases contributes to primary energy conservation through their function as redox-linked proton pumps²⁻⁵. With developments in protein structure and thermodynamic analysis and time-resolved high-resolution spectroscopy, the organization of the catalytic centres of the oxidase and the identity and structure of intermediates in the oxygen-reducing, proton-pumping catalytic cycle are emerging.

We have not attempted here to review this progress in detail. Instead, we have tried to integrate certain key findings, some of which have not been reconciled previously, in a new view of the O_2 reduction and proton translocation mechanism. The intimate interrelation between these two functions, and the central part played by the binuclear centre (Fig. 1) in both, are critical in understanding the molecular operation of this class of enzymes.

Cell respiration and the fitness of O_2

Water is exceedingly stable thermodynamically (hence its abundance) and photosynthetic organisms have developed well controlled mechanisms for its oxidation⁶. The waste in this process is gaseous O_2 . The evolutionary consequences of oxygenic photosynthesis were severe as a reducing atmosphere became oxidizing, and strategies for dealing with a new and reactive environment had to be developed. Unlike water, dioxygen is stabilized only kinetically and, if uncontrolled, it is thermodynamically capable of participating in catastrophic reactions that can lead to the disruption of cellular processes and cell death.

On the other hand, the appearance of O_2 in the atmosphere about 2×10^9 years ago must also have exerted a very strong evolutionary pressure towards organisms able to use O_2 as an oxidant of hydrocarbon foodstuffs. Much more energy could be released than was previously possible with other thermodynamically less effective oxidants. This evolutionary choice is evident today: all higher organisms are designed to capture atmospheric dioxygen, distribute it to the cells, and to reduce it to water in the process of cell respiration.

The fitness of oxygen to biology⁷ derives from two aspects of its structure and chemistry. First, there is a strong thermodynamic driving force for its reduction to water by biological reductants. Hence, it is ideal as a terminal oxidant as this maximizes the free energy made available to the cell. Secondly, the ground electronic state of atmospheric O_2 is a triplet state, having two unpaired electrons, which imposes spin restrictions on its reaction with singlet-state two-electron donors. One-electron reduction is sluggish as well, mainly because of a very large difference between the O-O bond energies in the O_2 molecule and the superoxide radical $O_2^{\cdot-}$. In essence, then, atmospheric O_2 is thermodynamically reactive with respect to reduction to water, but kinetically stabilized. It is therefore ideally suited to catalytic manipulation in enzyme active sites.

The poor chemical reactivity of dioxygen is important as terrestrial life would otherwise be jeopardized by spontaneous burning of organic material. The lethal potential of O_2 is evident under conditions in which its chemistry is uncontrolled. If the spin restriction is removed by spin-pairing, singlet oxygen (1O_2) is produced, which reacts rapidly and deleteriously in the cell. When O_2 is partially reduced, superoxide ($O_2^{\cdot-}$), hydrogen peroxide (H_2O_2) and hydroxyl radical ($OH^{\cdot}$) may be formed, all of which are implicated in oxygen toxicity⁸.

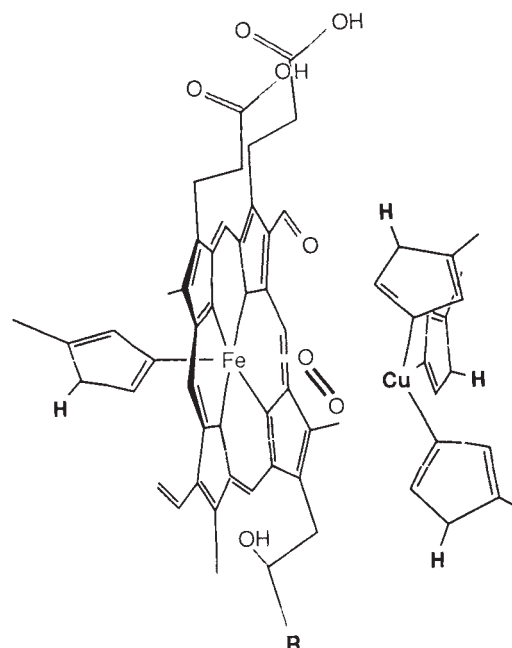


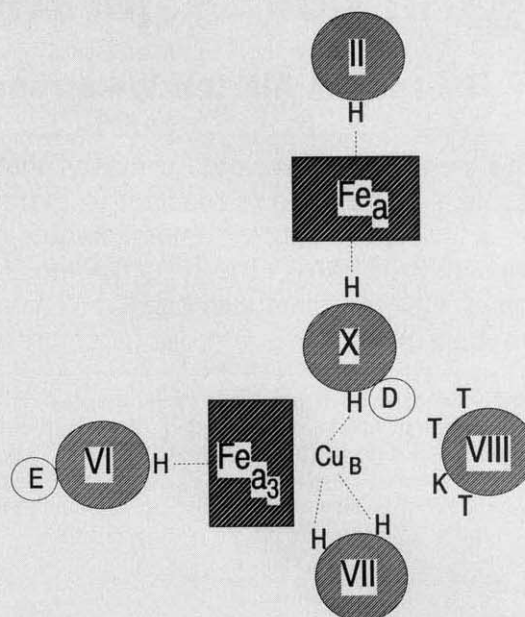
FIG. 1 Bimetallic haem iron-copper reaction centre in cytochrome oxidases with O_2 bound. The Fe-Cu distance is not known in this state, and may vary in different intermediates of the catalytic cycle.

BOX 1 The major subunits

SUBUNIT I of cytochrome *c* oxidases has 12 membrane-spanning helices. It shows primary structure homology to the quinol oxidases, although there are additional membrane-spanning helices in the latter (see refs 15, 19, 20); their numbering is such that any defined helix corresponds to the homologous one in all oxidases. Subunit I contains the binuclear centre as well as the low-spin haem. Six histidines, which are probable ligands to the haems and to Cu_B (Fig. 2), are conserved in all known cytochrome *c* and quinol oxidases. These histidines occur in helices II, VI, VII and X, and have been assigned as ligands to the respective metals (see figure) according to recent mutagenesis work²⁴⁻²⁶. The haem planes are perpendicular to the membrane plane, which suggests a structural motif where the haems are sandwiched between membrane-spanning helices. The Fe-Cu distance in the binuclear site is <5 Å in the oxidized enzyme, and the Fe-Fe distance between the haems is 15–20 Å (see ref. 27 and refs therein). The homology between subunits I from the different oxidases suggests that the geometries and locations of the redox centres are very similar in all these proteins. The figure shows a model of central aspects of subunit I, including the binuclear centre and the low-spin haem, depicted in a view perpendicular to the membrane plane (see also refs 24–26, 28). In this model it is speculated that the two haems (shaded) may be orientated with their planes perpendicular to one another, which would minimize direct electron transfer between them (see text). Five central membrane-spanning helices are indicated by Roman numerals. Some conserved amino acids in these helices, such as the metal-ligating histidines²⁴⁻²⁶ and two acidic residues (circled), which are discussed in the text, are indicated in one-letter code (D, aspartate; E, glutamate; H, histidine; K, lysine; T, threonine).

Subunit II traverses the membrane only twice, with the major C-terminal protein mass on the outside of the membrane. In some bacterial enzymes, a cytochrome *c* domain has been added to the C terminus¹⁵. Sequence analysis of subunit II from several species has suggested that the EPR-detectable Cu_A centre (Fig. 2) is bound to the non-helical C-terminal domain, which shows homology to blue copper proteins²⁹. Spectroscopic data³⁰ have established that two cysteines and at least one histidine are the Cu_A ligands. The only fully conserved cysteines in the terminal oxidases are those found in the Cu_A-binding domain of subunit II. It is possible that two, and not one, Cu ions are involved in the Cu_A site³¹. The quinol oxidases lack the Cu_A centre as well as the putative Cu_A ligands^{17-20,32}. In the cytochrome *c* oxidases, the Cu_A-binding domain also comprises several conserved acidic residues, which may contribute

to the cytochrome *c*-binding site³³. These are also lacking in the quinol oxidases.



Subunit III has five to seven transmembranous helices in the different oxidases. But it is lacking altogether from the *Sulfolobus* enzyme¹⁹. Subunit III contains no redox centre. Deletion of the corresponding gene in *P. denitrificans* gives an incompletely assembled enzyme, but the redox centres are formed³⁴. Subunit III was thought to play a part in the proton-pumping activity of the enzyme, but it can be dissociated from the protein with retention of at least some of this activity, and gene deletion and site-directed mutagenesis have shown that this subunit has no role in proton pumping³⁵. Subunit III may have a chaperonin-like function in enzyme assembly^{34,35}.

In the sometimes tumultuous history of the research on dioxygen metabolism (for reviews, see refs 9–12), notable developments were the elucidation of the cytochrome-containing respiratory chain, the realization that the terminal O₂-reducing catalyst is an iron-containing haem protein, and the implication from spectroscopic and magnetic studies that a binuclear haem *a*₃ iron-copper (Cu_B) centre is the dioxygen-reducing site in the terminal oxidase (Figs 1 and 2). By this time the terminal oxidase in the inner membrane of the mitochondrion had become known as cytochrome *c* oxidase, reflecting its immediate reductant, and a second haem (haem *a*) and a second copper (Cu_A) had been established as intermediary redox cofactors in the transfer of electrons from cytochrome *c* to the binuclear centre (Fig. 2a).

This enzyme, which is unique in eukaryotic cells, is a member of a family of related enzymes. Variants within this family are found in the cell membranes of aerobic bacteria¹³⁻²⁰. In certain thermophilic bacteria, for example, a *b*-type haem has replaced the low-spin haem *a* chromophore¹⁶. One subclass of the terminal oxidases uses a quinol rather than cytochrome *c* as the electron donor (Fig. 2b). Enzymes of this kind have been described in *Escherichia coli* (cytochrome *o*; refs 14, 15, 17), *Bacillus subtilis* (cytochrome *aa*₃-600; refs 18, 20) and *Sulfolobus acidocaldarius* (cytochrome *aa*₃; ref. 19). All these show clear structural and functional homology with the cytochrome *c* oxidases. Also, at least one of the haems of cytochrome *o* was a novel type, haem *O*, which is related to haem *A*, but with the formyl group replaced by a methyl residue^{21,22}. Despite the variations, both the binuclear centre²³ and the proton-pumping function⁵ in cytochrome *o* are retained. This has focused attention on the binuclear centre as the functional heart of this class of enzymes.

The protein framework that binds the metal centres in the family of terminal oxidases has emerged in the past 10 years. In the mammalian mitochondrial enzyme the three heaviest subunits (I–III) are encoded by mitochondrial DNA. A variety of smaller subunits are encoded in the nucleus. The various oxidases of aerobic bacteria are structurally much simpler, usually containing only the heavier subunits (I–III), which are homologous with their mitochondrial counterparts. There is now strong support for the view that subunits I and II form the catalytic core of all these enzymes, and details of the binding of the redox centres are emerging (Box 1).

Catalytic mechanism of O₂ activation

The chemical properties of dioxygen dictate some of the essential features that must characterize the catalytic binuclear site of the respiratory enzymes to facilitate its safe and efficient reduction to water. This site contains high-spin haem iron to overcome the spin restriction on O₂ reactivity. The occurrence of the redox-active copper in close apposition to the bound O₂ (Fig. 1) avoids the unfavourable one-electron reduction reaction. Moreover, heterolytic O–O bond scission can take place without the necessity of oxidizing the haem macrocycle or amino-acid side chains. This mechanistic feature sharply differentiates cytochrome and quinol oxidases from catalases, peroxidases and the cytochrome P-450 oxygenases. Partially reduced oxygen species do occur during catalysis, but rapid electron and proton transfer to them ensures that they are not released. At the same time, however, these electron and proton transfers must be carefully regulated so that the coupled proton translocation process is not short-circuited.

The redox potentials of the steps in the catalytic O₂ reduction

process are levelled relative to their solution values (Box 2). This helps the catalytic cycle kinetically and matches the thermodynamics of electron transfer to that of the linked proton-pumping events. Redox levelling, which also occurs in the photosynthetic water-oxidizing machinery³⁹, emerges as an important general principle of the way in which nature deals with multi-electron chemistry in enzyme active sites.

The two activities of the enzyme—O₂ reduction and proton translocation—are structurally, kinetically and energetically intertwined; together they dictate the dynamics of the catalytic cycle. An important general result from the studies discussed later is that electrons are transferred one at a time from the reducing substrate to the binuclear centre, where O₂ is bound. During catalysis, then, the binuclear centre cycles through intermediate states, which differ from one another with respect to valence, spin and ligation state of the metals. Each of the four electron transfers required to reduce O₂ to water has unique properties, and only two of them are linked to proton translocation³⁷. We focus first on the mechanism of reduction of dioxygen, where interpretation of kinetic data has perhaps not sufficiently taken into account the requirements associated with coupling to proton translocation. We then discuss the proton pumping action as such.

Catalysis by oxidase is remarkably rapid, maximal turnover numbers of the isolated enzyme in the steady state of more than 500 e⁻ per second are typical. Early on, the 'flow/flash' technique was developed (see Box 3), which, in combination with optical spectroscopy, can resolve individual electron transfer steps⁴⁰. Subsequently, this technique has yielded detailed kinetic and spectroscopic information on several intermediate species, beginning with the low-temperature triple-trapping studies of Chance *et al.*³⁸.

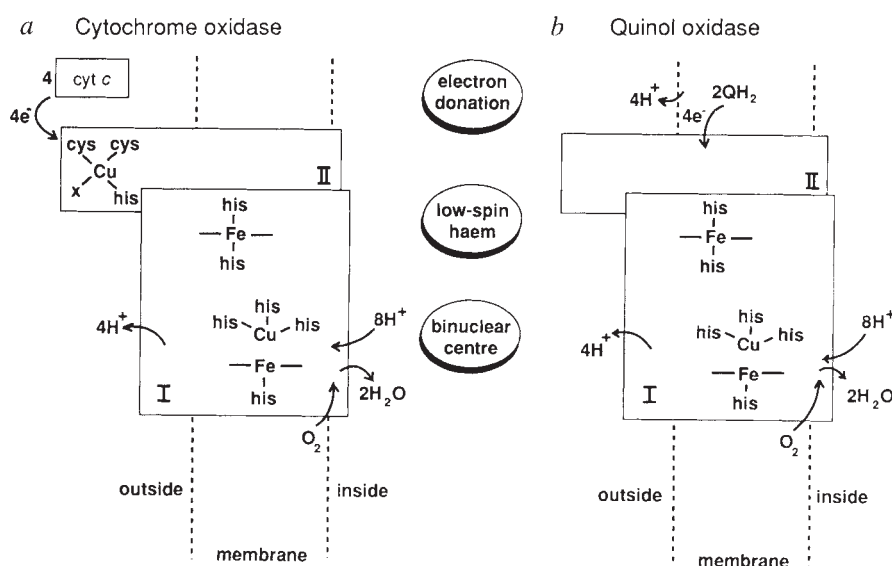
Another unique feature of the terminal oxidases—their redox-linked proton-pumping function—was used to reverse partially the catalytic cycle of O₂ reduction^{47,48}. This provided an independent means by which to explore individual catalytic steps and intermediates. This work has also provided insight into the redox potentials of these steps (ref. 36; Box 2), and their coupling to proton uptake in the formation of water from reduced oxygen⁴⁹. Some difficulty has arisen in integrating flow/flash and reversed electron transfer results into a single reaction mechanism, although the two approaches have gen-

erally yielded compatible data. In Fig. 3 we have attempted a synthesis of the results from both these approaches. This proposed mechanism expresses the view that the overall O₂ reduction reaction has two key features: (1) fast electron transfer to O₂ in the early stages of the reaction to trap the substrate in the active site as the peroxy species 3, and (2) slower subsequent electron transfers to bound, partially reduced oxygen intermediates that drive the proton translocation chemistry.

O₂ binding and trapping. The reduced binuclear centre 1 reacts with O₂ to produce the oxy adduct 2 (compound A; see Figs 1 and 3), which was originally inferred from optical measurements at low temperature³⁸. This species is formed transiently at room temperature^{50–52}, and time-resolved Raman spectroscopy was used to determine its structure^{53–55}. The Fe–O₂ stretching frequency is well reproduced by six-coordinate oxyhaem A models⁵⁶, which indicates that there is little interaction of the bound O₂ with Cu_B⁺. Chance *et al.*³⁸ noted that the O₂ binding in species 2 is fairly weak. The rapid electron transfers that produce the peroxy (P) species 3 are essential to trap the bound dioxygen, which raises the operational affinity of the site to that observed in steady-state kinetic measurements ($K_M \sim 10^{-7}$ M). The trapping chemistry seems to be optimized as O₂ does not bind to the ferrous haem until both electrons are present in the site^{57,58}.

The third electron problem. When the fully reduced enzyme reacts with O₂ in flow/flash experiments, the formation of species 2 is rapidly followed by electron transfer from haem *a* to the binuclear site, as deduced from electron paramagnetic resonance (EPR), optical and Raman measurements of haem *a* (refs 59–62); a rate constant of 2.5×10^4 s⁻¹ has been determined by these last two techniques at room temperature. This corresponds to the transition from 2 to 4 in Fig. 3. In fact, the intervening species 3 has not been observed in such experiments and hence its role as a catalytic intermediate, which was deduced from reversed electron transfer measurements^{36,47–49}, has been questioned⁶². It has been proposed that the primary intermediate—compound A—is already peroxidic (so equivalent to 3, not to 2) to account for an acceptor for the very fast electron transfer from haem *a* (refs 63, 64). But the Raman data discussed here are strong evidence against this possibility. The longer lifetime of 2 in the reaction of the mixed valence enzyme (haem *a*₃ and Cu_B reduced, haem *a* and Cu_A oxidized) with O₂ (refs 60, 65,

FIG. 2 Schematic representation of the arrangement of subunits I and II and the associated metal centres of the terminal respiratory enzymes: *a*, cytochrome oxidase and *b*, quinol oxidase. The dashed lines represent the membrane, with the inner and outer aqueous phases indicated. In the cytochrome oxidases, cytochrome *c* is the electron source, which reduces the Cu_A centre associated with subunit II; rapid intramolecular electron redistribution leads to the reduction of the low-spin haem *a* centre in subunit I. In the quinol oxidases, the reducing substrate is ubiquinol or menaquinol (QH₂). Because the Cu_A site is not present in this class of enzymes, the oxidation of QH₂ presumably occurs by electron transfer directly to the low-spin haem centre (but see ref. 19). In both types of enzyme, the low-spin haem serves as the electron-queueing point for the reduction of the binuclear centre. As discussed in detail in the text, the binuclear site is involved in both O₂ reduction and in redox-linked proton translocation. In these reactions, a total of eight protons are removed from the inner aqueous phase per O₂ reduced; four of these are translocated to the outer aqueous phase and four are consumed in the reduction of O₂ to two water molecules. Because QH₂ is a hydrogen atom carrier, whereas cytochrome *c* is a simple one-electron donor, the outside proton-to-electron

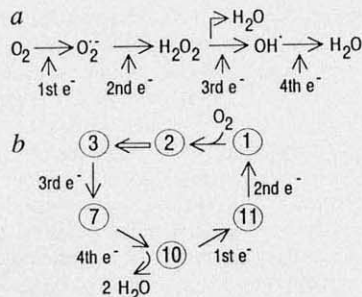


stoichiometry differs for the two classes of enzymes. For cytochrome oxidase this ratio is one; for quinol oxidase it is two.

BOX 2 Free energies of individual steps in the reduction of O₂ in aqueous solution and in the cytochrome oxidase reaction

THE thermodynamics of O₂ reduction in aqueous solution is characterized by different driving forces for the individual electron transfers: the conversion of O₂ to superoxide is unfavourable, but the reduction of hydroxyl radical is strongly driven. This is unsuitable for a fast enzymatic reaction or a reaction coupled to efficient proton translocation. In the enzyme the thermodynamics of the individual electron transfers are sharply modified and levelled relative to the electrochemical reaction. These two modes of O₂ reduction are compared here.

The schemes show *a*, the one-electron pathway of O₂ reduction to water via superoxide, peroxide and hydroxyl radical and *b*, principal intermediates of catalytic O₂ reduction by cytochrome oxidase, numbered as in Fig. 3.

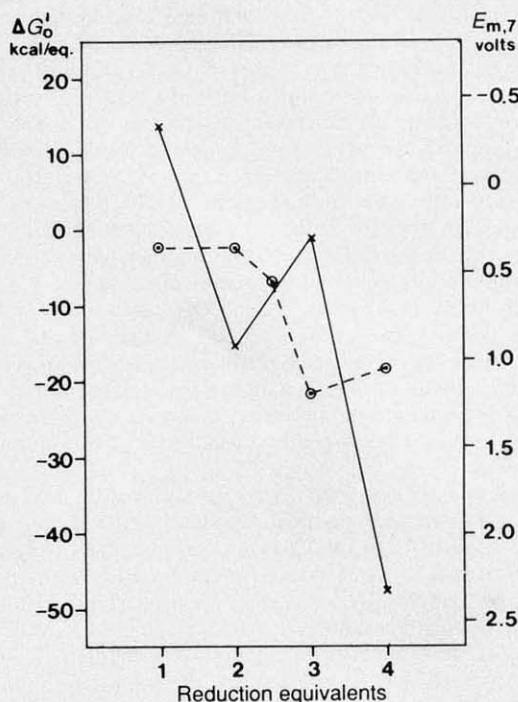


The corresponding $E_{m,7}$ values for *a*, O₂ reduction to water in aqueous solution (standard states: pH 7; fugacity of O₂, 1 atm; activity of other species except water, 1 molal), and *b*, for the redox couples, at zero protonmotive force, involved in the first, second, third and fourth electron transfers to the binuclear site of cytochrome *c* oxidase are shown below (see ref. 36 for the calculation procedure; the values in ref. 36 are corrected here, owing to the translocation of 2H⁺ per step in the third and fourth electron transfers; see ref. 37 and Fig. 3).

Step	<i>a</i> $E_{m,7}$ (volts)	Step	<i>b</i> $E_{m,7}$ (volts)
1st e ⁻	-0.33	10 → 11	+0.35
2nd e ⁻	+0.94	11 → 1	+0.35
		1 → 3	+0.29*
3rd e ⁻	+0.305	3 → 7	+1.20
4th e ⁻	+2.33	7 → 10	+1.05

* Obtained by difference as sum of the cycle is $\text{O}_2 + 4e^- + 4\text{H}^+ \rightarrow 2\text{H}_2\text{O}$; $E_{m,7} = 0.81 \text{ V}$

The graph compares the standard free energies for the four steps of dioxygen reduction to water, in aqueous solution (crosses) and in the enzyme (circles), with ferrocyanide *c* ($E_{m,7} = 260 \text{ mV}$) as the reductant.



Binding of O₂ to the reduced binuclear site 1, and trapping of the O₂ by internal electron transfer to form the peroxy state 3, is not coupled to electron or proton translocation. The reaction 1 → 3 is, therefore, not balanced by the protonmotive force or any fraction thereof. Together these steps occur with a standard free energy change of about -6 kcal mol⁻¹, shown as the extra step at 2.5 equivalents. As O₂ binding *per se* is nearly reversible³⁸, this free energy change is attributable mainly to the O₂-trapping step 2 → 3, which provides necessary irreversibility to the catalytic cycle at high protonmotive force.

66) has also been used as an argument against viewing 3 as a true intermediate⁶². Nonetheless, there are experimental observations that strongly support the occurrence of the peroxy species 3 in the catalytic cycle: when the mixed valence enzyme reacts with O₂, 3 is detected as a long-lived intermediate³⁸ and it is readily formed from 10 by reversed electron transfer^{36,47-49}.

These observations reflect a discrepancy in the interpretation of results from flow/flash and reversed electron transfer experiments. We believe the reason for this may be attributed to the linkage of dioxygen reduction to proton translocation, and specifically to the necessity of controlling the rate of both electron and proton delivery to the peroxy species 3 in this process. We will attempt to show that, if this linkage is sufficiently taken into account, the apparent discrepancy can be reconciled.

Role of redox cooperativity. The negative redox cooperativity in cytochrome oxidase has been known for some time¹², but its functional significance has not been understood. It is most pronounced between haem *a* on one hand, and haem *a*₃ and Cu_B on the other. It implies, for example, that reduction of Cu_B thermodynamically disfavors reduction of haem *a*, and vice versa. We suggest that this negative redox cooperativity is a crucial feature of the coupling of electron transfer to proton translocation. It may be specifically required to prevent electronic short-circuiting, that is, electron transfer uncoupled from proton translocation (see later). One of its consequences is that haem *a* and Cu_B are unlikely to be in their reduced state

simultaneously. Indeed, under steady-state turnover conditions, haem *a* is largely oxidized and O₂ reacts for the most part with the enzyme in what is effectively a mixed valence state (see ref. 67 and refs therein). Thus, the fully reduced enzyme is hardly present during aerobic turnover; it is only formed when the constraints on the metal valence imposed by negative cooperativity have been overcome under special conditions.

Consequently, we suggest that the rapid conversion of 2 into 4 is a step unique to the fully reduced enzyme. It represents premature reduction of the binuclear centre, and in the parlance of a proton pump, it corresponds to a 'slip' in the pumping cycle. In this interpretation, the failure to observe proton pumping in the early stages of the reduction of O₂ by the fully reduced enzyme⁶⁸ is a direct manifestation of this short-circuiting reaction. Thus the negative redox cooperativity may be designed specifically to prevent such short-circuited electron transfer from haem *a*. In this regard the oxidases are analogous to the cytochromes P-450, where rate control is necessary to avoid short-circuiting the oxygen bond insertion chemistry^{69,70}.

Scission on the O-O bond. As with the terminal oxidases, peroxy species occur as precursors to dioxygen bond cleavage in the cytochrome P-450 oxygenases and in a variety of peroxidases and catalases^{69,70,71}. In these systems, however, oxygen activation proceeds by a route distinctly different from that in the terminal oxidases. In cytochrome P-450 it is helped by the strongly electron-donating proximal thiol ligand to the haem

iron; O-O bond scission produces the activated ferryl iron species (Fe^{4+}) and an oxidized haem macrocycle. The role of proximal ligand basicity in O-O bond cleavage is also apparent in the peroxidases and catalases. Here, the occurrence of an endogenous acid function in the distal pocket seems to be essential as well⁶⁹. As with the cytochromes P-450, dioxygen bond cleavage produces Fe^{4+} and a radical species—either the oxidized haem macrocycle or an amino-acid side chain.

The means by which O-O bond scission occurs in the terminal oxidases is clearly different. The haem a_3 iron-histidine bond-stretching frequency in the reduced enzyme indicates that the imidazole moiety is neutral and not strongly electron-donating⁷². The formation of 3 is not accompanied by proton uptake⁶⁸ and its long lifetime in the mixed-valence enzyme suggests that the acid/base chemistry necessary to initiate cleavage of the O-O bond is not inherent to the site. Rather, conditions that facilitate this chemistry appear to be set up in the binuclear centre after the peroxy species is formed, and require the electron and proton transfer reactions that produce 4 and 5 in Fig. 3. The presence of the second metal, Cu_B , obviates the need to oxidize organic moieties during dioxygen bond cleavage, further distinguishing the respiratory oxidases from the catalase and peroxidase enzymes. Thus, O-O bond cleavage in Fig. 3 is suggested first to require reduction of Cu_B , then substrate protonation. The end-bound peroxide structures for intermediates 3, 4 and 5 in this process are suggested by time-resolved Raman measurements^{73,74}. Peroxy structures in which the ligand is bridged to Cu_B or occurs as a side-bound species^{75,76} are less likely in view of inferences drawn from ligand-binding studies⁷⁷ and the alkyl peroxide data discussed below.

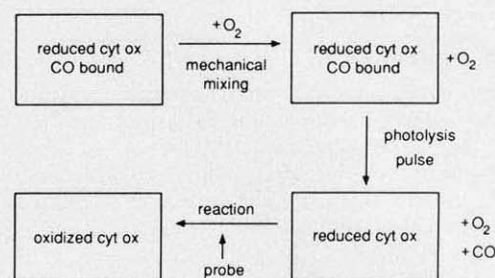
Net uptake of two protons is coupled to the transition of 3 to 7 (ref. 36). The rate constant for the uptake of the first of these protons has been determined⁶⁸, suggesting that the reaction $4 \rightarrow 5$ was monitored. The possibility that this reaction reflects a Bohr protonation cannot, of course, be discounted. Low-temperature work showing an EPR signal from Cu_B led to the postulate of a bridged ferrous/cupric peroxide species, shown here as 6, which precedes rupture of the O-O bond⁷⁸. A bridged structure such as 6 might provide a means for O-O bond cleavage that complements the part played by the distal histidine in the peroxidase active site. On the other hand, the data in ref. 78 do not necessitate such bridging, and the rupture of the O-O bond may instead be facilitated by uptake of a second proton⁷⁷. The decay of 6 into the ferryl species 7 is entropy-driven, consistent with O-O bond cleavage at this step⁷⁸. EPR data on Cu_B from experiments with isotopically labelled O_2 (ref. 79), together with the proton uptake stoichiometry in the transition of 3 to 7 (ref. 36), allow us to place H_2O derived from O_2 in the coordination sphere of Cu_B in the latter intermediate (Fig. 3).

The ferryl (F) intermediate 7 was originally observed in reversed electron-transfer experiments⁴⁷, where it was shown to result from one-electron oxidation of the ferric/cupric centre 10, an assignment supported by low-temperature trapping⁷⁸. This intermediate was observed in optical flow/flash measurements at room-temperature⁵², and the $\text{Fe}^{4+}=\text{O}$ structure has since been confirmed by time-resolved Raman experiments (refs 61, 80, 81, but see ref. 73). The frequency of the iron-oxygen stretch is similar to that seen with a variety of peroxidase compounds II, for which the ferryl structure is well established (see ref. 56 and refs therein).

The fourth electron. Injection of the fourth electron into the binuclear site initiates the second cycle of proton pumping, and involves the transition from ferryl 7 to the fully oxidized form of the site, 10, through a hydroxide-bound form, 9 (ref. 36). The occurrence of a ferric hydroxide intermediate has been verified by Raman spectroscopy⁸¹. Flow/flash measurements, in which proton extrusion from vesicular preparations of the enzyme was monitored, show that the transition from 7 to 10 is coupled to proton translocation, even with the fully reduced enzyme⁶⁸, in agreement with reversed electron-transfer results³⁷.

BOX 3 The flow/flash technique: converting a thermal reaction into a photochemically initiated reaction

THE reaction between reduced cytochrome oxidase and dioxygen is complete in a few milliseconds, which is comparable to the mixing times of conventional stopped-flow mixers and thus limits the usefulness of this technique in studying the oxidase/ O_2 reaction. Gibson and Greenwood⁴⁰ took advantage of the kinetic parameters (see table below) of cytochrome oxidase to develop the flow/flash technique, which overcomes this limitation. In this approach, the reduced CO-bound enzyme is photolysed in the presence of O_2 , and the ensuing electron transfers, dioxygen bond cleavage chemistry, and proton movements can be monitored with time resolution that is limited only by the duration of the photolysis pulse.



The method relies on rapid dioxygen binding juxtaposed with slow CO rebinding kinetics. The data in the table show that of the oxygen-metabolizing haem proteins, the terminal oxidases are most suitable for investigation with this technique. The flow/flash approach has been adapted to a variety of spectroscopic methods, and the results from these investigations are described in the text.

Rate and equilibrium parameters for the reaction of CO and O_2 with several haem-containing systems

	$k_{\text{on}}(\text{CO})$ ($\text{M}^{-1} \text{s}^{-1}$)	$k_{\text{off}}(\text{CO})$ (s^{-1})	$K_{\text{eq}}(\text{CO})$ (M^{-1})	$k_{\text{on}}(\text{O}_2)$ ($\text{M}^{-1} \text{s}^{-1}$)
Cytochrome oxidase	7×10^4	0.023	3×10^6	1×10^8
P-450 _{cam} + cam	3.7×10^4	0.14	2.6×10^5	7.7×10^5
Cytochrome c peroxidase	2×10^3	0.0092	2.2×10^5	4.5×10^4
Model haem	8.2×10^6	0.025	3.3×10^8	3.5×10^7
R-state haemoglobin				
α chains	6.5×10^6	0.01	6.5×10^8	5.9×10^7
Myoglobin (horse)	5×10^5	0.17	2.9×10^6	1.4×10^7
Leghaemoglobin	1.3×10^7	0.016	7.9×10^8	1.2×10^8

Data collected from refs 41 and 42. The model compound is protohaem mono-3-(1-imidazolyl) propylamide monomethyl ester in toluene (O_2 data) or benzene (CO data).

The high time resolution afforded by flash photolysis of CO has also been put to use in another context, in studies of the dynamics of the binuclear centre after CO photolysis. Two questions, in particular, have been addressed: (1) the relaxation time of the haem a_3 pocket following photolysis, and (2) the fate of the departing CO. These studies have also provided information about the ligand-binding process. Infrared spectroscopy has shown that CO ligates Cu_B after photolysis at low temperature⁴³. The haem a_3 iron-histidine bond relaxes to its equilibrium geometry on the microsecond timescale after CO photolysis at room temperature⁴⁴. The migration of the CO to the Cu_B site occurs in picoseconds at room temperature⁴⁵. The Cu_B^+-CO species formed has a lifetime of $\sim 1 \mu\text{s}$ before the CO is released into bulk solution. From data accumulated with several time-resolved spectroscopic techniques, it has been proposed that the a_3/Cu_B pair has a ligand shuttle activity, and that Cu_B may be the initial binding site of the incoming O_2 . Transient binding of O_2 to Cu_B^+ has been proposed⁴⁶. Transfer of O_2 to its chemically reactive haem a_3 site occurs in microseconds, whereupon the reaction sequence indicated in Fig. 3 is initiated. The proposal of Cu_B as a waystop for O_2 on its route to the haem a_3 iron has interesting mechanistic implications and is likely to be the focus of much further research.

The peroxide reaction. Figure 3 suggests that H_2O_2 may react with the ferric/cupric state of the binuclear centre 9 and/or 10 and short-circuit catalysis by producing the peroxy species 3, and upon addition of electrons, the ferryl 7 and full oxidized 10 states. The general consensus (see for example refs 9, 82, 83) is that species with optical characteristics of the peroxy and ferryl intermediates can be produced this way, although the method usually yields these species in less than stoichiometric amounts. Nevertheless, this effect of H_2O_2 supports the assignment of intermediate 3 as a ferric-cupric peroxidic species. But it is unlikely that the O_2^{2-} bridges the metals in this compound, because organic peroxides can also react with 10 to form a peroxy intermediate with spectral characteristics indistinguishable from 3 (ref. 84). This favours the single-ended bonding structure shown in Fig. 3.

Proton translocation

Protonic coupling is widely recognized as the universal principle of primary energy conservation in respiration^{1,85}. Yet the discovery that cytochrome oxidase functions as a proton pump² gained acceptance only after several years of controversy (see for example, ref. 86, p. 48). The molecular mechanism of proton pumping by the terminal oxidases remains one of the main challenges in this area.

Previously it was thought that the proton-pumping action of cytochrome oxidase was coupled to each one of the four electron transfers from cytochrome *c* to the binuclear site. Haem *a* and Cu_A have both been proposed as candidates in linking redox transitions to proton movement (see ref. 87 for a review). These ideas have recently been jeopardized by two independent findings. First, it was shown in reverse electron flow experiments that translocation of two protons is coupled to the $3 \rightarrow 7$ (peroxy to ferryl) transition and that an additional two protons are translocated upon the $7 \rightarrow 10$ (ferryl to oxidized) transition³⁷. Proton translocation therefore occurs asymmetrically in the catalytic cycle: the first two electron transfers to the binuclear centre are not coupled to proton pumping; the third and fourth result in translocation of two protons each. Proton pumping is thus coupled to electron transfer only when oxygen is ligated at the binuclear centre.

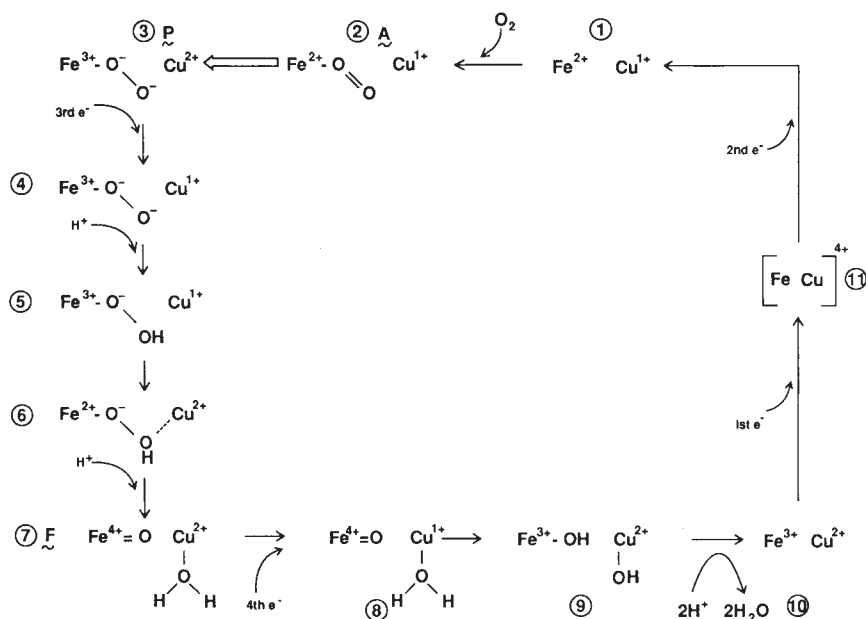
These experiments have also yielded information on the ener-

getics of the O_2 reduction cycle (ref. 36; Box 2). At present, only standard free energies can be estimated for the major reaction steps at zero protonmotive force. These must be interpreted with caution as the relative occupancies of the reaction intermediates are not known in the steady state. For kinetic reasons, however, the ratio of oxidized to reduced forms for any of the redox couples is unlikely to be offset by more than two orders of magnitude (120 mV or 2.8 kcal) in the steady state; accordingly, the standard free energies should show the principal thermodynamic trends in the system. According to this analysis (Box 2), the conversion of the oxy intermediate 2 to the peroxy form 3 provides the physiologically important irreversibility of respiration at a high-energy state of the cell. This is consistent with the conclusion that this step is essential in trapping the bound O_2 in the catalytic site.

The second observation involves the quinol oxidase, cytochrome *a*, which was shown also to sustain redox-linked proton translocation⁵. Owing to the structural homology to the cytochrome *c* oxidases, proton pumping in the quinol oxidase almost certainly proceeds by the same or a very similar mechanism. As the quinol oxidases contain neither the Cu_A centre nor the formyl group of the haem A chromophore, mechanisms that involve either of these in proton translocation become very unlikely.

The role of Cu_B . In Fig. 3 we take the view that the third and fourth electrons are accepted by Cu_B . This is crucial for our proposed role of Cu_B in energy conservation, and is based on evidence suggesting that Cu_B provides the electron entry point into the binuclear centre (see for example ref. 88). The proton-pumping transitions ($3 \rightarrow 7$ and $7 \rightarrow 10$) include these electron entry reactions, as well as the subsequent reactions within the binuclear centre to produce 7 and 10, respectively (Fig. 3). Therefore, the proton-pumping events could be coupled to the electron transfers from cytochrome *c* (through Cu_A and haem *a*) to Cu_B , and need not necessarily be linked to the subsequent chemistry at the binuclear site. But there are several reasons to favour reactions at the binuclear centre as central in the proton-pumping mechanism. The $E_{m,7}$ of Cu_B is not higher than ~ 0.4 V (see refs 12, 23). Unless this value has been drastically raised in intermediates 3 and 7, the electron transfer into Cu_B is not *per se* sufficiently exergonic to drive proton translocation. The subsequent exergonic reactions at the binuclear site could, of

FIG. 3 Proposed mechanism for dioxygen activation and reduction in cytochrome and quinol oxidases. The binuclear centre, which consists of haem iron proximally ligated by histidine, and the histidine-ligated Cu_B centre (see Figs 1 and 2), is represented here by 'Fe Cu' for clarity. The oxygen activation sequence begins with the reduced centre 1, which binds dioxygen to produce the oxy species 2 (compound A). Intracentre electron transfer traps dioxygen as the peroxy (P) species 3. The thermodynamic irreversibility of this step is represented here by the heavier arrow for the $2 \rightarrow 3$ transition. Input of the third electron to 3 from the electron-queueing low-spin haem centre generates 4. After protonation (5) and intracentre electron redistribution (6), oxygen-oxygen bond scission produces the ferryl (F) species 7. In the steady state, translocation of two protons is coupled to the $3 \rightarrow 7$ transition (not shown); in flow/flash experiments with the fully reduced enzyme, electronic short-circuiting causes the fast transition from 2 to 4 that by-passes proton-translocating transitions in Cu_B and uncouples proton translocation normally associated with the $3 \rightarrow 7$ sequence (see text). Input of the fourth electron to 7 from the low-spin haem centre generates 8. After electron and proton redistribution, the hydroxy species 9 is formed. Subsequent protonation releases water and generates the oxidized enzyme 10. In the sequence $7 \rightarrow 10$, an additional two protons are translocated. The cycle is reset with the input of the first and second electrons. The one-electron



reduced intermediate in this process, 11, is here shown with the electron distributed over the centre; O_2 binding in this state is inhibited and occurs only when the fully reduced binuclear centre 1 is regenerated.

course, provide the necessary driving force by poisoning Cu_B at a potential substantially (about 300 mV) higher than $E_{m,7}$. But this would require a ratio of 10^5 between the concentrations of oxidized and reduced forms of Cu_B in the steady state. This is improbable for kinetic reasons, in that a rate constant of $\sim 10^9 \text{ s}^{-1}$ would then be required for the oxidation of Cu_B .

Secondly, it is reasonable to assume that electron transfer from cytochrome *c* to Cu_B occurs by the same pathway, irrespective of the state of the binuclear site in the catalytic cycle (but see ref. 87 and below for a possible exception). Branched electron transfer that may bypass Cu_A (refs 59, 60, 78) is no longer probable. Recent data^{89,90} suggest a linear input pathway in cytochrome oxidase (Fig. 2), in which Cu_A is the immediate oxidant of cytochrome *c*, the former then equilibrating rapidly with haem *a* (ref. 91). As proton pumping is associated with the steps involving reduction of 3 and 7, but not with reduction of 10 to 1, it follows that proton pumping must be linked to the chemistry triggered by the individual electron transfers to the binuclear site.

We conclude that electron transfer within the binuclear centre, from Cu_B to the Fe-oxygen system, including the subsequent chemistry of the latter, provides the major driving force for proton pumping. The mechanism of proton pumping is probably the same when linked to the sequences 3 → 7 and 7 → 10 (Fig. 3) because it is unlikely that the enzyme employs two different mechanisms for the same purpose. But as conversion of the peroxy state 3 to the ferryl state 7 is chemically very different from the reduction of the ferryl iron in 7 to ferric in 10, proton-pumping is unlikely directly to involve the oxygen chemistry or the haem a_3 site. The peroxy and ferryl structures merely provide thermodynamically efficient acceptors of the electrons from Cu_B , which is common to both reactions. Cu_B itself is hence the most probable redox element of the proton pump.

Functioning of the proton pump. There have been detailed discussions on the properties and requirements of the redox element of a proton pump^{12,92-95}. Although these have not revealed anything about the actual molecular mechanism, they may be helpful in rationalizing kinetic and thermodynamic findings. Figure 4a shows the now commonly used cubic scheme of H^+ pumping, where the redox element (Cu_B) must cycle between discrete electronic and protonic input and output states. For example, if input electron transfer from haem *a* could occur to Cu_B^{2+} while the latter is still in its output state, electron short-circuiting would abolish proton pumping (Fig. 4c). A control function must minimize the probability of this if proton pumping is to occur efficiently. As discussed above, the negative redox cooperativity provides for such control. When Cu_B^{1+} occurs in the output state (Fig. 4), the input electron donor (the low-spin haem) normally remains oxidized as the result of this interaction. Therefore, after fast electron transfer from Cu_B^{2+} to the Fe-oxygen system in the output configuration, Cu_B^{2+} has time to revert to the input state. But this may not be the case if Cu_B and haem *a* are in the reduced state simultaneously, as in the fully reduced enzyme. Then fast electron transfer from haem *a* to the binuclear site is observed immediately after the binding of O_2 , as discussed in the preceding section. This short-circuiting reaction is depicted in Fig. 4c. It emerges from this scenario that the control of electron transfer from the low-spin haem to Cu_B is exerted thermodynamically, that is, by the negative redox cooperativity, which makes short-circuiting combinations of redox states improbable.

A second level of control is also evident in the terminal oxidases. In contrast to the cytochromes P-450, peroxidases and catalases, conversion of the peroxy to the ferryl species in the oxidases is controlled by electron and proton transfers into the binuclear centre, not by the internal structure and chemistry of the O_2 -binding site itself. This catalytic feature complements the negative redox cooperativity and is essential in allowing the input/output relaxation and subsequent reduction of Cu_B that

couple proton translocation to the conversion of 3 into 7 in Fig. 3.

It is interesting that the fast oxidation of haem *a* and the failure to observe intermediate 3 with the fully reduced enzyme, juxtaposed with the long lifetime of 3 in the mixed-valence enzyme, and the negative redox cooperativity, all find a common explanation with the postulated role of Cu_B in proton pumping. This view also readily explains why only substoichiometric proton translocation has been observed in the reaction of the fully reduced enzyme with O_2 (ref. 68).

In Fig. 3, transfer of the first two electrons to the binuclear site completes the catalytic cycle and 1 is regenerated. These electron transfers are not coupled to proton translocation. We need to enquire how this 'decoupling' is achieved mechanistically, because it is not explained satisfactorily by the lower driving force. Two major possibilities emerge, both of which depend on the fact that the structure of the binuclear centre is considerably different during the 10 → 1 and 2 → 10 transitions. In 10 and 11 the haem a_3 iron is in a high-spin state, unliganded to oxygen, and at least in 10 the Fe_{a3} is $< 5 \text{ \AA}$ from Cu_B . In contrast, after reduction of 10 to 1, and after binding of O_2 , which precedes the proton-translocating events, the Fe- Cu_B distance is enhanced beyond $5 \text{ \AA}$ (refs 9, 96) and the haem iron

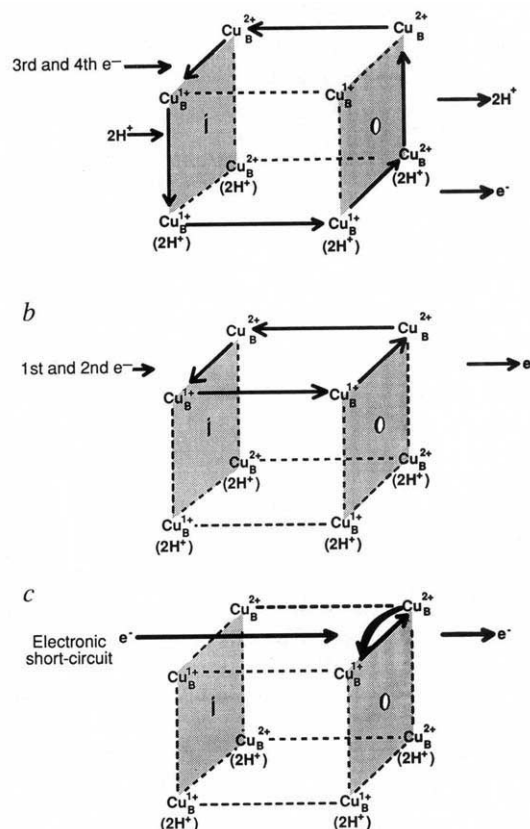


FIG. 4 Cubic scheme of the functioning of Cu_B as the redox element of the proton pump (see also refs 12, 92, 94). The shaded squares *i* and *o* represent the four electronic and protonic input and output states, respectively. Electronic input is from the low-spin haem *a*; output is to the haem a_3 /oxygen site. Protonation of a site(s) that is coupled to the redox centre is shown under Cu_B in parentheses; this notation does not necessarily imply that the protonation reactions occur at sites close to the copper. In *a*, the complete proton-pumping cycle, which is linked to the third and fourth electron transfer events in the enzyme (see text), is outlined by arrows. In *b*, the sequence of events during transfer of the first and second electron is outlined by arrows. In this pathway, which is favoured over *a*, when the haem a_3 centre is high-spin and unliganded to oxygen, the linkage to proton translocation does not occur. *c*, The proposed 'short-circuited' pathway of the third electron under conditions in which the fully reduced enzyme reacts with O_2 . The fully reduced enzyme is unusual under physiological conditions, but occurs in flow/flash-type experiments (see the text for details).

assumes a low-spin state with an oxygenous ligand. Therefore, decoupling of the $10 \rightarrow 1$ transition from proton pumping could simply result from a structural modulation of electron transfer pathway, such that the low-spin haem a is able to donate electrons directly to haem a_3 , whereby the Cu_B is short-circuited. In the model of Box 1 this could occur by a small change in the haem-haem angle, making it more favourable for direct electron transfer²⁸. The second possibility requires that electron transfer through the Cu_B centre can occur in modes both coupled to and decoupled from proton pumping. To achieve this, the structural change in the binuclear site referred to above may modulate the rate constant of the input to output transition for the unprotonated state relative to the rate of protonation of the input state of the proton pump (compare Fig. 4a and b).

This analysis stresses a probable role of Cu_B in energy conservation by the terminal oxidases, but it also strongly emphasizes the necessity of cooperativity between Cu_B and its immediate electron donor, the low-spin haem, to make the energy conservation process effective. In this view, the low-spin haem performs an essential electron-queueing function that ensures an orderly input of electrons to the binuclear centre. We think that it is significant that not only the binuclear site, but also its strong interaction with the low-spin haem, have been conserved in proton-pumping oxidases of both aa_3 and o type. On the other hand, the electron donor (cytochrome c or quinol) and the electron input site of the enzyme (Cu_A in cytochrome c oxidases) are variable; they are apparently not mechanistically essential for proton translocation.

H⁺ transfer mechanism. With respect to molecular mechanisms, it may be timely to consider whether there is any analogy to the mechanism of proton translocation by bacteriorhodopsin, the light-driven proton pump of *Halobacterium halobium*. Here membrane-embedded aspartic residues plus a protonated Schiff's base of the chromophore have been directly implicated in proton transfer (see ref. 97 and refs therein). In the subunits I of cytochrome c oxidases there are two conserved acidic residues predicted to be close to the binuclear site. One of them is not conserved in cytochrome o (an aspartic residue in helix X, see Box 1; ref. 15) and only one more fully conserved acidic residue is predicted to reside in the membrane plane (a glutamic acid in helix VI, see Box 1; ref. 15). This residue is not conserved in the enzyme from *Sulfolobus*¹⁹, and its mutation to glutamine in cytochrome o does not impair proton translocation in *E. coli* sphaeroplasts (J. Thomas and A. Puustinen, unpublished). Directed mutagenesis of the conserved glutamic acid residue in subunit III, which is the site for dicyclohexyl carbodiimide-binding, has shown that this residue is not involved in proton pumping³⁵. Thus, at present a close analogy to the bacteriorhodopsin mechanism does not seem obvious.

Our proposal of Cu_B as the redox element of the proton pump implies that the states through which it cycles during catalysis are linked in a controlled fashion to input/output states of proton-binding site(s), as indicated schematically in Fig. 4. At present we know nothing about the nature of these sites, nor about their mode of linkage to Cu_B . They may occur at a long distance from Cu_B ; in contrast, a relatively direct coupling mechanism might involve the metal ligands themselves^{28,98,99}. In the latter case, conserved threonine, serine and lysine residues in helix VIII of subunit I (Box 1) could be involved in proton transfer, possibly interacting with ligands of Cu_B (compare ref. 28). Alternatively, the proton-translocating events might occur at a considerable distance from the redox centre and its coordination sphere, with indirect coupling through the protein structure (compare the Bohr effect in haemoglobin and the helix-twist model¹⁰⁰).

Epilogue

We have described the substantial progress that has been made in understanding the chemistry of O_2 reduction. The proton translocation function is not yet understood, although the past

few years have provided further insight. Generally, the research on cellular respiration is still hampered by the lack of good diffracting crystals of any one of the interrelated terminal oxidases. Three-dimensional structural information will forward the field and provide the necessary framework for concrete molecular interpretations of kinetic data. But the kinetic approach will, in any case, continue to be indispensable for any detailed understanding of function. □

Gerald T. Babcock is in the Department of Chemistry, Michigan State University, East Lansing, Michigan 48824, USA, and Mårten Wikström is in the Department of Medical Chemistry, University of Helsinki, Helsinki 000170, Finland.

- Mitchell, P. *Eur. J. Biochem.* **95**, 1-20 (1979).
- Wikström, M. *Nature* **266**, 271-273 (1977).
- Soloz, M., Carafoli, E. & Ludwig, B. *J. Biol. Chem.* **257**, 1579-1582 (1982).
- Sone, N. & Hinkle, P. *J. Biol. Chem.* **257**, 12600-12604 (1982).
- Puustinen, A., Finel, M., Virkki, M. & Wikström, M. *FEBS Lett.* **249**, 163-167 (1989).
- Babcock, G. T. *et al. Biochemistry* **28**, 9557-9565 (1989).
- George, P. in *Oxidases and Related Redox Systems* Vol. 1 (eds King, T. E., Mason, H. S. & Morrison, M.) 3-33 (Wiley, New York, 1965).
- Rodgers, M. A. J. & Powers, E. L. (eds) *Oxygen and Oxy-Radicals in Chemistry and Biology* (Academic, New York, 1981).
- Naqui, A., Chance, B. & Cadenas, E. A. *Rev. Biochem.* **55**, 137-166 (1986).
- Beinert, H. *Chemica Scripta* **28A**, 35-40 (1988).
- Malmström, B. G. *Chem. Rev.* **90**, 1247-1260 (1990).
- Wikström, M., Krab, K. & Saraste, M. *Cytochrome Oxidase—A Synthesis* (Academic, New York and London, 1981).
- Ludwig, B. *FEMS Microbiol. Rev.* **46**, 41-56 (1987).
- Anraku, Y. A. *Rev. Biochem.* **57**, 101-132 (1988).
- Saraste, M. *Q. Rev. Biophys.* **23**, 331-336 (1990).
- Zimmerman, B. H., Nitsche, C. I., Fee, J. A., Rusnak, F. & Munk, E. *Proc. natn. Acad. Sci. U.S.A.* **85**, 5779-5783 (1988).
- Chepur, V., Lemieux, L., Au, D. C.-T. & Gennis, R. B. *J. Biol. Chem.* **265**, 11185-11192 (1990).
- Lauraeus, M., Haltia, T., Saraste, M. & Wikström, M. *Eur. J. Biochem.* **197**, 699-705 (1991).
- Lübben, M., Kolmer, B. & Saraste, M. *EMBO J.* **11**, 805-812 (1992).
- Santana, M. *et al. J. Biol. Chem.* (in the press).
- Puustinen, A. & Wikström, M. *Proc. natn. Acad. Sci. U.S.A.* **88**, 6122-6126 (1991).
- Wu, W. *et al. J. Am. Chem. Soc.* **114**, 1182-1187 (1992).
- Salerno, J. C., Bolgiano, B., Poole, R. B. & Ingledew, W. J. *J. Biol. Chem.* **265**, 4364-4368 (1990).
- Lemieux, L. J., Calhoun, M. W., Thomas, Ingledew, W. J. & Gennis, R. B. *J. Biol. Chem.* (in the press).
- Minagawa, J., Mogi, T., Gennis, R. B. & Anraku, Y. *J. Biol. Chem.* (in the press).
- Shapleigh, J. *et al. Proc. natn. Acad. Sci. U.S.A.* (in the press).
- Holm, L., Saraste, M. & Wikström, M. *EMBO J.* **6**, 2819-2823 (1987).
- Haltia, T. & Wikström, M. in (ed. Ernster, L.) *Molecular Mechanisms in Bioenergetics* (Elsevier, Amsterdam, in the press).
- Steffens, G. J. & Buse, G. *Hoppe-Seyler's Z. physiol. Chem.* **360**, 613-619 (1979).
- Martin, C. T., Scholes, C. P. & Chan, S. I. *J. Biol. Chem.* **263**, 8420-8429 (1988).
- Kroneck, P. M. H., Antholine, W. A., Riest, J. & Zumft, W. G. *FEBS Lett.* **242**, 70-74 (1988).
- Puustinen, A., Finel, M., Haltia, T., Gennis, R. B. & Wikström, M. *Biochemistry* **30**, 3936-3942 (1991).
- Capaldi, R. A. *Archs Biochem. Biophys.* **280**, 252-262 (1990).
- Haltia, T. *et al. EMBO J.* **8**, 3571-3579 (1989).
- Haltia, T., Saraste, M. & Wikström, M. *EMBO J.* **10**, 2015-2021 (1991).
- Wikström, M. *Chemica Scripta* **28A**, 71-74 (1988).
- Wikström, M. *Nature* **338**, 776-778 (1989).
- Chance, B., Saronio, C. & Leigh, J. S. Jr *J. Biol. Chem.* **250**, 9226-9237 (1975).
- Bouges-Bocquet, B. *Biochim. biophys. Acta* **594**, 85-103 (1980).
- Gibson, Q. & Greenwood, C. *Biochem. J.* **86**, 541-554 (1963).
- Mims, M. P., Porras, A. G., Olson, J. S., Noble, R. W. & Peterson, J. A. *J. Biol. Chem.* **258**, 14219-14232 (1983).
- Chang, C. K. & Dolphin, D. *Proc. natn. Acad. Sci. U.S.A.* **73**, 3338-3342 (1976).
- Alben, J. O., Moh, P. P., Flamingo, F. G. & Altschuld, R. A. *Proc. natn. Acad. Sci. U.S.A.* **78**, 234-237 (1981).
- Findsen, E. W., Centeno, J., Babcock, G. T. & Ondrias, M. R. *J. Am. Chem. Soc.* **109**, 5367-5372 (1987).
- Woodruff, W. H. *et al. Proc. natn. Acad. Sci. U.S.A.* **88**, 2588-2592 (1991).
- Blackmore, R. S., Greenwood, C. & Gibson, Q. H. *J. Biol. Chem.* **266**, 19245-19249 (1991).
- Wikström, M. *Proc. natn. Acad. Sci. U.S.A.* **78**, 4051-4054 (1981).
- Wikström, M. *Chemica Scripta* **27B**, 53-58 (1987).
- Wikström, M. *FEBS Lett.* **231**, 247-252 (1988).
- Babcock, G. T., Jean, J. M., Johnston, L. N., Palmer, G. & Woodruff, W. H. *J. Am. Chem. Soc.* **106**, 8305-8306 (1984).
- Orii, Y. *J. Biol. Chem.* **259**, 7187-7190 (1984).
- Orii, Y. *Ann. N.Y. Acad. Sci.* **550**, 105-117 (1988).
- Varotsis, C., Woodruff, W. H. & Babcock, G. T. *J. Am. Chem. Soc.* **111**, 6439-6440 (1989); **112**, 1297 (1990).
- Ogura, T., Takahashi, S., Shinzawa-Itoh, K., Yoshikawa, S. & Kitagawa, T. *J. Am. Chem. Soc.* **112**, 5630-5631 (1990).
- Han, S., Ching, Y.-C. & Rousseau, D. L. *Proc. natn. Acad. Sci. U.S.A.* **87**, 2491-2495 (1990).
- Oertling, W. A., Kean, R. T., Wever, R. & Babcock, G. T. *Inorg. Chem.* **29**, 2633-2645 (1990).
- Lindsay, J. G., Owen, C. S. & Wilson, D. F. *Archs. Biochem. Biophys.* **169**, 492-505 (1975).
- Malatesta, F., Sarti, P., Antonini, G., Vallone, B. & Brunori, M. *Proc. natn. Acad. Sci. U.S.A.* **87**, 7410-7413 (1990).
- Clare, G. M., Andréasson, L.-E., Karlsson, B. & Malmström, B. G. *Biochem. J.* **185**, 139-154 (1980).
- Hill, B. C., Greenwood, C. & Nicholls, P. *Biochim. biophys. Acta* **853**, 91-113 (1986).
- Varotsis, C. & Babcock, G. T. *Biochemistry* **29**, 7357-7362 (1990).
- Han, S., Ching, Y.-C. & Rousseau, D. L. *Proc. natn. Acad. Sci. U.S.A.* **87**, 8408-8412 (1990).
- Wilson, M. T., Jensen, P., Aasa, R., Malmström, B. G. & Vänngård, T. *Biochem. J.* **203**, 483-492 (1982).
- Malmström, B. G. A. *Rev. Biochem.* **51**, 21-59 (1982).
- Varotsis, C., Woodruff, W. H. & Babcock, G. T. *J. Biol. Chem.* **265**, 11131-11136 (1990).

66. Han, S., Ching, Y.-C. & Rousseau, D. L. *Biochemistry* **29**, 1380–1384 (1990).
67. Morgan, J. E. & Wikström, M. *Biochemistry* **30**, 948–958 (1991).
68. Oliveberg, M., Hallén, S. & Nilsson, T. *Biochemistry* **30**, 436–440 (1991).
69. Poulos, T. L. *Adv. Inorg. Biochem.* **7**, 1–52 (1987).
70. Tabushi, I. *Coord. Chem. Revs.* **86**, 1–42 (1988).
71. Traylor, T. G. *Pure appl. Chem.* **63**, 265–274 (1991).
72. Babcock, G. T. in *Biological Applications of Resonance Raman Spectroscopy* Vol. 3 (ed. Sprio, T. G.) 294–346 (Wiley, New York, 1988).
73. Ogura, T., Takahashi, S., Shinzawa-Itoh, K., Yoshikawa, S. & Kitagawa, T. *Bull. Chem. Soc. Jpn.* (in the press).
74. Varotsis, C., Zhang, Y. & Babcock, G. T. *Proc. natn. Acad. Sci. U.S.A.* (submitted).
75. Ahmad, S. *et al. Inorg. Chem.* **27**, 2230–2233 (1988).
76. Proniewicz, L. M., Paeng, I. R. & Nakamoto, K. *J. Am. Chem. Soc.* **113**, 3294–3303 (1991).
77. Yoshikawa, S. & Caughey, W. S. *J. biol. Chem.* **265**, 7945–7958 (1990).
78. Blair, D. F., Witt, S. N. & Chan, S. I. *J. Am. Chem. Soc.* **107**, 7389–7399 (1985).
79. Hansson, O., Karlsson, B., Aasa, R., Vänngård, T. & Malmström, B. G. *EMBO J.* **1**, 1295–1297 (1982).
80. Ogura, T., Takahashi, S., Shinzawa-Itoh, K., Yoshikawa, S. & Kitagawa, T. *J. biol. Chem.* **265**, 14721–14723 (1990).
81. Han, S., Ching, Y.-C. & Rousseau, D. L. *Nature* **348**, 89–90 (1990).
82. Wrigglesworth, J. M. *Biochem. J.* **217**, 715–719 (1984).
83. Vygodina, T. & Konstantinov, A. A. *Biochem. biophys. Acta* **973**, 390–398 (1989).
84. Vygodina, T. V., Schmidmeyer, K. & Konstantinov, A. A. *Biol. Membr.* (in Russian) (in the press).
85. Williams, R. J. P. *J. theor. Biol.* **1**, 1–17 (1961).
86. Mitchell, P. in *Of Oxygen, Fuels and Living Matter* (ed. Semenza, G.) 1–160 (Wiley, New York, 1981).
87. Chan, S. I. & Li, P. M. *Biochemistry* **29**, 1–12 (1990).
88. Nicholls, P. & Wrigglesworth, J. M. *Ann. N.Y. Acad. Sci.* **550**, 59–67 (1988).
89. Kobayashi, K., Uhe, H. & Hayashi, K. *J. biol. Chem.* **264**, 7976–7980 (1989).
90. Hill, B. C. *J. biol. Chem.* **266**, 2219–2226 (1991).
91. Antalis, T. M. & Palmer, G. *J. biol. Chem.* **257**, 6194–6206 (1982).
92. Wikström, M. & Krab, K. *Biochim. biophys. Acta* **549**, 177–222 (1979).
93. Blair, D. F., Gelles, J. & Chan, S. I. *Biophys. J.* **50**, 713–733 (1986).
94. Krab, K. & Wikström, M. *Biochim. biophys. Acta* **895**, 25–39 (1987).
95. Malmström, B. G. *FEBS Lett.* **250**, 9–21 (1989).
96. Scott, R. A. *Rev. Biophys. biophys. Chem.* **18**, 137–158 (1989).
97. Henderson, R. *et al. J. molec. Biol.* **213**, 899–929 (1990).
98. Wikström, M. & Krab, K. in *Energy Conservation in Biological Membranes* (eds Schäfer, G. & Klingenberg, M.) 128–139 (Springer, Berlin and Heidelberg, 1978).
99. Mitchell, P. *Ann. N.Y. Acad. Sci.* **550**, 185–198 (1988).
100. Williams, R. J. P. *FEBS Lett.* **226**, 1–7 (1987).

ACKNOWLEDGEMENTS. We thank T. Haltia, J. E. Morgan and C. Varotsis for discussion, and E. Schmidt and H. Vuorenmäki for help with the figures. Supported by the US National Institutes of Health (G.T.B.) and the Sigrid Jusélius Foundation (M.W.).

ARTICLES

Anisotropy of the inner core from differential travel times of the phases PKP and PKIKP

Kenneth C. Creager

Geophysics Program, University of Washington, Seattle, Washington 98195, USA

The difference between the travel times of compressional waves that turn in the Earth's liquid outer core (PKP-BC) and those that run in the solid inner core (PKIKP-DF) is primarily sensitive to structure near the inner-core boundary. Such differential travel times measured from short-period waveforms are consistent with small-amplitude deviations from radial symmetry except when ray paths are nearly parallel to the Earth's spin axis. These near-axial paths consistently produce large travel-time anomalies, providing compelling evidence for large-amplitude, axisymmetric anisotropy in the outermost portions of the inner core.

THE eigenfrequencies of the Earth's normal modes that are sensitive to seismic velocity structure at and below the core-mantle boundary are split much more than expected from the Earth's rotation and ellipticity. This anomalous splitting was first measured by Masters and Gilbert¹ and has since been thoroughly analysed^{2–8}. There is general agreement on the measurements themselves, and on the inference that the structure causing the splitting is symmetric with respect to the Earth's spin axis (axisymmetric) and is at or below the core-mantle boundary. Two classes of models have been proposed to explain these observations. One is aspherical seismic-velocity structure dominated by the zonal spherical harmonic of degree 2 within the outer core, which is fast at the poles^{2,3}. The second is axisymmetric anisotropy of the inner core with the fast direction parallel to the Earth's spin axis^{4–7}.

The travel-time residuals of rays turning in the inner core (PKIKP or P_{DF}) reported by the International Seismological center (ISC) also show a strong axisymmetric pattern which was interpreted by Poupinet *et al.*⁹ as zonal, degree-2 topography of the inner core, and by Morelli *et al.*¹⁰ and Shearer *et al.*¹¹ as inner-core anisotropy. There is some uncertainty, however, in

the interpretation of absolute travel times, owing to potential bias from earthquake mislocation and unknown aspherical structure of the crust, mantle and core-mantle boundary. Shearer and Toy¹² recognized that, in principle, the structure of the inner core can be resolved by analysing differential travel times of hand-picked PKP-BC (P_{BC}) and P_{DF} phases. The paths for these two rays remain within 300 km of each other throughout the mantle, but the DF phase enters the inner core whereas BC turns near the base of the outer core (Fig. 1), so differential travel-time residuals are most plausibly explained by structure either in a boundary layer at the base of the outer core, or structure in the inner core.

Like Shearer and Toy¹², I have picked differential times of P_{BC} minus P_{DF} from short-period recordings of the Global Digital Seismograph Network (GDSN). Among the common seismograms, our results are nearly identical, and show small, but geographically consistent, residuals with a root-mean square of 0.4 s. But to resolve axisymmetric inner-core anisotropy one needs observations of rays at a range of directions through the inner core relative to the Earth's spin axis. Whereas Shearer and Toy¹² analysed only one seismogram corresponding to a ray direction within 40° of this axis, I analyse 20 and find that

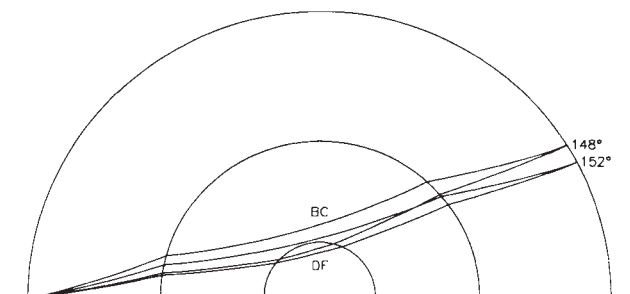


FIG. 1 Ray paths of P_{DF} entering the solid inner core and P_{BC} turning in the liquid outer core from a surface-focus earthquake to stations at epicentral distances of 148° and 152°.

TABLE 1 Models of inner-core anisotropy

Model	B (km s ⁻¹)	C (km s ⁻¹)	δv^* (%)	Data	Phase	Distance range	Ref.
1			3.5 [†]	IDA	modes		4
2			0.8–1.2 [‡]	IDA	modes		7
3	-0.350	0.700	0.0–3.2 [§]	ISC	DF	170° < Δ < 180°	10
4			1.0	ISC	DF	170° < Δ < 180°	10
5	-0.092	0.200	1.0	ISC	DF	120° < Δ < 140°, 154° < Δ < 180°	11
6	0.012	0.056	0.6	ISC	BC-DF	145° < Δ < 155°	12
7	-0.165	0.557	3.5	GDSN	BC-DF	146° < Δ < 160°	this work

* Polar velocity minus equatorial velocity, divided by equatorial velocity.

[†] Calculated equatorial PKIKP travel time minus polar time, divided by equatorial time, for rays straight through the inner core.

[‡] Calculated equatorial PKIKP travel time minus polar time, divided by equatorial time, for rays turning at all depths in the inner core.

[§] δv varies as $(r/R_{IC})^2$, where R_{IC} is the inner-core radius.

18 of them have large positive residuals in the range +2 to +4 s; these are 5 to 11 standard deviations away from the normal population of 244 observations whose ray angle with respect to the Earth's spin axis exceeds 40°. Because of their dependence on ray direction, and not on geographical sampling location, these new observations are most plausibly explained by anisotropy. The ray-path sampling of the differential times allows us to infer that the anomalous structure is at the base of the outer core, or at the top of the inner core. Finally, the sign

of the anomalous travel-time residuals is consistent with the sign of the anomalous eigenfrequency splitting if the structure is assumed to be in the inner core, but has the opposite sign for an explanation in terms of structure at the base of the outer core. This study supports previous models of inner-core anisotropy except that, owing to the new observations, the amplitude of inner-core anisotropy is about three times larger than inferred from most previous body-wave estimates.

The amplitude ratios (BC/DF) of the 18 anomalous seismograms are also anomalous, with DF being consistently three times smaller than normal. The very small amplitudes of DF may explain why Shearer and Toy analysed only one of these anomalous seismograms, and may contribute to a systematic bias in the ISC Bulletin times, leading other researchers to underestimate the size of anisotropy in the inner core.

Travel-time and amplitude observations

I extracted 2,500 short- and intermediate-period seismograms of P'_{BC} and P'_{DF} recorded in the epicentral distance range 146° to 160° on the GDSN from 1980 to 1986. These data were distributed by the National Earthquake Information Center. Of these seismograms 264 produce very clear arrivals of both P'_{BC} and P'_{DF} . Seismograms in which depth phases may have interfered with the later phase (P'_{BC}) were discarded. The waveform shapes of P'_{BC} and P'_{DF} are typically very similar, although the amplitude ratio P'_{BC}/P'_{DF} varies from 0.5 to 30.0, and P'_{DF} is typically, but not always, relatively depleted in high frequencies. Differential travel times were measured from the corresponding peaks and troughs of the two phases. These times were then averaged with weight given to the first cycle of each arrival. Because I only analysed the best data, and was able to match the shape of the two waveforms, the observed differential times are rarely, if ever, off by one or more cycles. Dominant periods are typically ~1 s. About 5% of the observations have very clear impulsive arrivals for both phases, but BC is dominated by high frequencies, whereas the high frequencies in DF are strongly attenuated. In these cases the differential times correspond to the difference between first arrivals. Only 13 of the times were read from intermediate-period records, and in every case the times are consistent with short-period records in similar geographical source-receiver geometries.

Differential travel-time residuals were calculated with respect to isotropic PREM¹³ at a period of 1 s, using the earthquake locations given by the National Earthquake Information Center. Times were corrected for ellipticity¹⁴. Beyond 152°, the BC branch diffracts around the inner core. Travel times for diffracted waves are calculated by linear extrapolation of the curve of travel time against distance for BC turning rays. Positive residuals represent early DF or late BC, and correspond to fast velocities in the inner core of slow velocities at the base of the outer core.

Epicycle mislocations of 20 km and depth mislocations of 100 km cause errors in the differential times ranging from 0.15 to 0.30 s depending on the depth and epicentral distance. Subduction-zone earthquakes can be systematically mislocated by

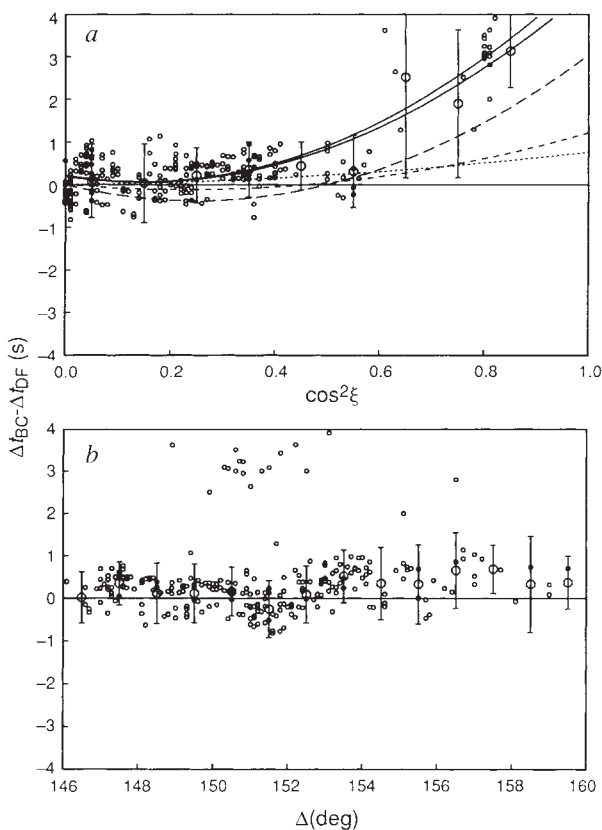


FIG. 2 Hand-picked differential travel-time residuals ($P'_{BC} - P'_{DF}$) plotted with respect to $\cos^2 \xi$ (a) and epicentral distance (b). ξ is the angle between the Earth's spin axis and the inner-core leg of the ray path. Bars show mean and two standard deviations of data binned at (a) 0.1 and (b) 1.0° intervals. Statistics in lower diagram are calculated using only times with $\cos^2 \xi < 0.6$. In a, differential travel times are calculated for surface-focus events at an epicentral distance of 150° using the anisotropy models of refs 10 (long-dashed line), 11 (short-dashed line) and 12 (dotted line) and that developed here (lower solid line). The latter model is also calculated for an epicentral distance of 152° (upper solid line). Models 3, 5, 6 and 7 of Table 1 are displayed.

20 km or more^{15,16}, causing systematic errors which would appear as geographically consistent signals of a few tenths of a second, but would rarely cause errors in excess of 0.5 s.

The histogram of differential travel-time anomalies has a remarkable distribution. Most of the times (92%) have a near-gaussian distribution with a mean of 0.2 s and a standard deviation of 0.4 s. The remaining 8% are clustered with a mean of 2.9 s, a full seven standard deviations away from the majority of the observations. The anomalous arrivals occur over a broad range of epicentral distances (Fig. 2b), event depths from 0 to 170 km and event magnitudes from 5.5 to 6.3, but the plot of residuals against the angle ξ between the Earth's spin axis and the ray path through the inner core cleanly separates all the anomalous phases from the normal phases (Fig. 2a).

Figure 3 shows six examples of anomalous arrivals. Some of my differential travel-time measurements could be in error by a few tenths of a second, but the very early P'_{DF} arrivals are unmistakable. Note also that the BC/DF amplitude ratios vary from 10 to 20. The amplitude ratios for the 'normal' observations in the same epicentral distance range cluster around 5. The BC/DF amplitude ratios for all paths with $\cos^2 \xi > 0.7$ are consistently three times larger than for paths to the same epicentral distances, but with $\cos^2 \xi < 0.6$. The small amplitude of the DF phase for large values of $\cos^2 \xi$ makes the observations of 'anomalous' phases difficult to obtain because DF is seldom above the noise level.

Anisotropy or heterogeneity?

The cause of these anomalous observations can best be discerned by analysing the ray-path sampling of the Earth's interior. Figure 4 shows the observed anomalies, plotted in map view at the turning points, the inner-core boundary intersection points of DF, and the source and receiver points. The turning-point plot shows remarkable coherence, even though the r.m.s. signal in the 'normal' observations is only 0.4 s. This consistency suggests that the r.m.s. reading error is much less than 0.4 s, perhaps as small as 0.2 s. The anomalous times (large plus signs) are widely distributed. There are 14 under Columbia, and one under each of western China, the Philippines, the south central Pacific and the central Atlantic. With the exception of the 'anomalous'

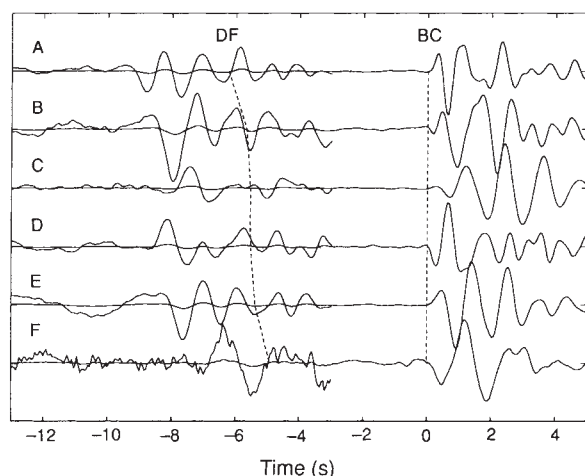


FIG. 3 Six vertical-component, short-period seismograms aligned on the P'_{BC} arrival (time=0). Dashed lines indicate the differential travel times predicted by Earth model PREM. From -13 to -3 s each seismogram is displayed twice, once at the same scale as the BC phase, and once magnified by a factor of 10 so that the DF phase can be seen. For these anomalous paths ($\cos^2 \xi \approx 0.8$) the BC-DF differential time is, on average, 3 s greater than predicted by PREM. Seismograms A-E are recorded at station COL (Alaska) from earthquakes under the South Sandwich Islands which are separated by as much as 130 km in epicentre and vary in depth from 30 to 170 km. F is recorded at RSNT (Canada) from an earthquake under the Atlantic-Indian Rise.

phases, the times are dominated by negative residuals in the western hemisphere and positive in the eastern hemisphere. These small, but consistent, residuals are plausibly explained by large-scale aspherical structure (perhaps degree 1) of the inner core, or the base of the outer core, with small (<0.5%) velocity contrast. Alternatively, they may be caused by systematic biases in earthquake mislocations.

The other two diagrams of Fig. 4 are more complicated and internally less consistent. In particular, it is worth noting that in the plots of both turning points and inner-core intersection points, the western hemisphere mostly contains negative residuals, except for the 'anomalous' (plus) times, most of which are also in this hemisphere. Geographically, the anomalous phases are widely distributed and interspersed with the normal observations, suggesting that they cannot be explained by a model of large-scale isotropic heterogeneity.

Alternatively, we can check whether the observations are consistent with simple models of anisotropy. Figure 5a displays the differential time residuals in an equal-area, azimuthal projection of the P'_{DF} ray direction in the inner core. The centre of the diagram corresponds to rays travelling parallel to the spin axis, and the perimeter represent rays in the equatorial plane. This diagram separates the anomalous observations from the normal ones. It also separates the normal observations into groups with considerable internal consistency.

Following refs 10-12, I consider simple models of anisotropy. Let us consider uniform anisotropy of the inner core with cylindrical symmetry about the Earth's spin axis. For weak anisotropy, the compressional wave speed can be approximated¹² by $\delta v_p(\xi) = A' + B' \cos 2\xi + C' \cos 4\xi$ or equivalently by¹⁰ $\delta v_p(\xi) = V_{eq}(1 + \epsilon \cos^2 \xi + \sigma \cos^2 \xi \sin^2 \xi)$ where ξ is the angle between the spin axis and the P'_{DF} ray segment through the inner core. For the PREM velocity model, rays through the inner core follow circular paths whose radii of curvature are so large that a ray changes direction by less than 1° during transit through the inner core. To a good approximation, each ray can be characterized by a constant angle ξ . The above equations can be rewritten in a third equivalent form as a parabola in $\cos^2 \xi$:

$$\delta v_p(\cos^2 \xi) = A + B \cos^2 \xi + C (\cos^2 \xi)^2 \quad (1)$$

using the conversion equations $A = A' - B' + C' = V_{eq}$, $B = 2B' - 8C' = V_{eq}(\epsilon + \sigma)$, $C = 8C' = -V_{eq}\sigma$. Using this parameterization, and considering δv_p to be anomalies with respect to PREM, I have inverted the hand-picked differential travel times using least squares and obtain $A = 0.018 \text{ km s}^{-1}$, $B = -0.165 \text{ km s}^{-1}$, and $C = 0.557 \text{ km s}^{-1}$. The wave speed in the polar direction is 3.5% faster than in the equatorial direction.

This simple two-parameter model explains 69% of the variance in the 264 observations. To examine whether the data require an axisymmetric structure, I consider cylindrically symmetric anisotropy, but vary the direction of the symmetry axis. For each axis, I invert for the best model parameters and calculate the variance reduction (Fig. 5b). The best fit is achieved when the symmetry axis is 5° from the spin axis at 85° S, 300° E, but acceptable models (variance reductions of 50% or more) are achieved for models in which the symmetry axis is displaced up to 5° from the South Pole along a longitude of 180° and 35° along the 0° longitude (Fig. 5b). It is clear from the ray-angle diagram (Fig. 5a) that more observations with ray angles near the spin axis will help constrain the models, especially observations with longitudes near 270°.

Comparison with previous models

Table 1 summarizes models of inner-core anisotropy. Models 3-7 are parameterized in the same way as above (equation (1)), except that the amplitude of the anisotropy in model 3 varies as r^2 , whereas models 4-7 assume uniform anisotropy. Theoretical travel-time anomalies depend only on ξ , on epicentral distance (Δ) and weakly on hypocentral depth. Because the ray paths are nearly straight, and the inner-core velocity varies

slowly with depth, the theoretical time residuals can be approximated by

$$\delta t^{\text{DF}}(\Delta, \cos^2 \xi) \approx (-T(\Delta)/V)\delta v_p(\cos^2 \xi) \quad (2)$$

where $T(\Delta)$ is the travel time through the inner core and $V = 11 \text{ km s}^{-1}$ is the velocity near the top of inner core. Using this scaling, I compare the actual observations with the travel times predicted at $\Delta = 150^\circ$ ($T = 124 \text{ s}$) for models 3, 5, 6 and 7 (Fig. 2a). The solid lines show the weak dependence on Δ by displaying predicted times for $\Delta = 150$ and 152° , the distance range of the bulk of our 'anomalous' observations (Fig. 2b). Model 3 is corrected for the r^2 dependence by multiplying the right-hand side of equation (2) by $1/3 + 2/3(r_t/R_{\text{IC}})^2$ where r_t is the turning radius of the ray and R_{IC} is the radius of the inner core. For $\Delta = 150^\circ$, this factor is 0.77.

All seven models are consistent with the fast direction being parallel to the spin axis, but there is considerable variation in the amplitude of the anisotropy. Morelli *et al.*¹⁰ proposed two models of anisotropy consistent with absolute P'_{DF} times from the ISC in the epicentral distance range 170 to 180° . These rays propagate nearly straight through the centre of the inner core and thus sample nearly all depths of the inner core. Their data could be explained by anisotropy decaying as r^2 with 3.2% contrast at the surface of the inner core (model 3), or by uniform anisotropy with a 1% contrast (model 4). An analysis of absolute P'_{DF} times from the ISC in the epicentral distance ranges 120 – 140° and 154 – 180° , which correspond to rays turning within 70 km of the inner-core boundary and more than 310 km below

this boundary, respectively, is consistent with 1% uniform anisotropy (model 5)¹¹. Anomalous times reported here occur in an intermediate distance range (148 – 156° , Fig. 2b), but the apparent anisotropy is 3–4 times larger in amplitude.

The waveform data set analysed here is similar to that of Shearer and Toy¹², and the observed times are nearly identical (compare Fig. 4a with their Fig. 6) except that they measured only one time with $\cos^2 \xi > 0.6$. Their model matches both their observations and the subset of mine with $\cos^2 \xi < 0.6$, but is not consistent with my observations for $\cos^2 \xi > 0.6$ (Fig. 2a). Shearer and Toy¹² also analysed arrival times reported in the ISC Bulletin. They discarded differential times (BC – DF) with residuals in excess of 1.5 s , apparently because the scatter in these times is small (see their Fig. 11) and their hand-picked residuals never exceeded $\pm 1 \text{ s}$. They found that weak (0.6%) anisotropy is consistent with both the ISC and GDSN observations. I have analysed the ISC data in the epicentral distance range 150 – 156° , but with a residual cut-off of 3.5 s . There are 95 observations with $\cos^2 \xi > 0.6$, 53 of which corresponded to paths from 25 earthquakes in the South Sandwich Islands to nine stations in Alaska. The geographical positions of sources and receivers are similar to those for seismograms A–E (Fig. 3). The mean ($\pm 1 \text{ s.d.}$) for these 53 observations is $2.6 \pm 0.8 \text{ s}$, consistent with the anomalous times measured here and consistent with 3% anisotropy. These times were rejected, however, by Shearer and Toy¹². No other geographical region is as well sampled, and all other regions show much less internal consistency. The mean residual of the remaining 42 times having $\cos^2 \xi > 0.6$ is 0.5 s . These small residuals are inconsistent with the anisotropy model presented here.

For comparison with body-wave models, I estimate δv (defined in Table 1) from mode-derived models as the theoretical PKIKP travel-time difference between pole-perpendicular and pole-parallel paths divided by the time of the pole-perpendicular paths. On the basis of modal data, Woodhouse *et al.*⁴ constructed model 1 with $\delta v > 3\%$ for rays through the centre of the Earth, whereas Li *et al.*⁷ produced model 2 with δv varying from 0.8 to 1.2% for different ray turning depths (estimated from their Fig. 13). The latter model is consistent with PKIKP travel times predicted by body-wave models 4–6. The argument consistently delivered in refs 4–7 is that there exists a model of axisymmetric anisotropy of the inner core that is consistent with the mode data, but that these data cannot uniquely determine inner-core anisotropy. Given the variety of published models that are consistent with the modes, it would seem that there exists a model of axisymmetric inner-core anisotropy of the form analysed by Li *et al.*⁷ that is consistent both with the modes and with the travel times presented here. Note, however, that the interpretation that mode splitting is primarily caused by inner-core anisotropy remains controversial. Observation of modes that had not previously been analysed, combined with an analysis of the radial distribution of elastic energy for these modes, has led Widmer⁸ to conclude that the primary cause for anomalous splitting is structure above the inner core, but below the mantle.

The hand-picked times analysed here, and a re-evaluation of some of the ISC differential times, suggest that previous body-wave studies have underestimated the size of the inner-core anomaly by between 40% (model 3) and a factor of 3 or more (models 4–6). The difference results primarily from new waveform observations of highly anomalous differential times in the critical ray-sampling direction nearly parallel to the Earth's spin axis. Observations for this sampling direction in the ISC catalogue are rare, but show consistently anomalous times for South Sandwich Islands earthquakes to Alaska at an epicentral distance of ~ 150 – 154° (this work), and in the distance range 170 – 180° (ref. 7). Other geographical or epicentral distance ranges do not show as strong an anomaly in the ISC catalogue¹¹. It is difficult to reconcile all these observations with a simple model of the inner core, although a model with large-amplitude

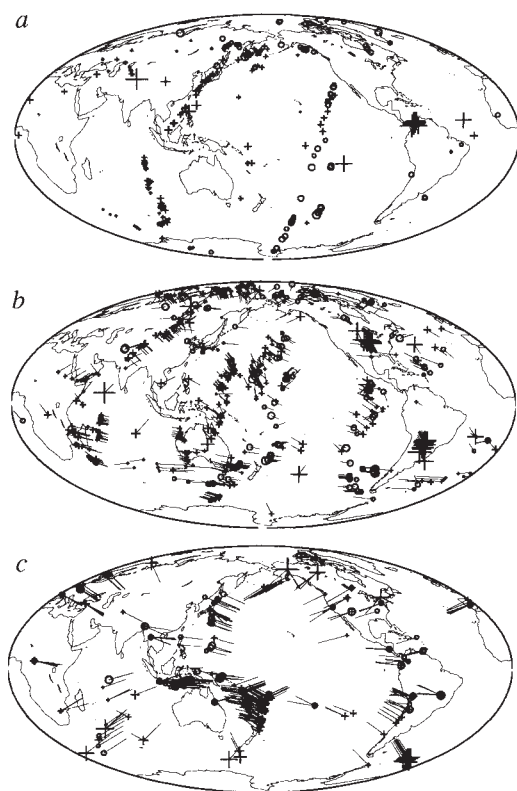
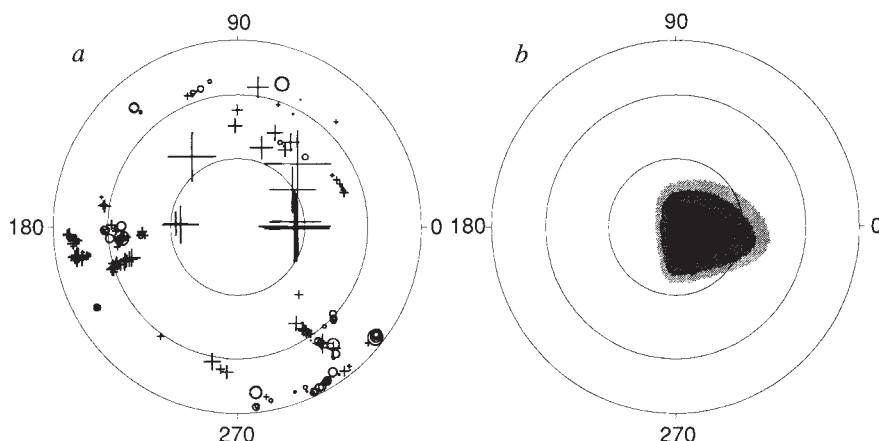


FIG. 4 Differential travel-time residuals ($P'_{\text{BC}} - P'_{\text{DF}}$) plotted as a function of (a) ray turning point (b) inner-core intersection points and (c) source and receiver locations. +, positive; O, negative. Residual magnitude is proportional to symbol size with the largest symbols representing a 4-s residual. In the two lower panels each observation corresponds to two geographical positions; line segments point towards the other position of the pair. The anomalous times, interpreted here as being caused by anisotropy, are the large plus symbols. In the upper diagram, there are 14 under Columbia, and one under each of western China, the Philippines, the south central Pacific and the central Atlantic.

FIG. 5 *a*, Differential travel-time residuals as a function of inner-core P'_{DF} ray direction displayed on an equal-area, azimuthal projection of the Southern Hemisphere. Ray direction latitude is 90° S at the centre and 0° at the circumference, with latitudes of 0 , 30 and 60° S shown by concentric circles. The longitude of the ray direction in the Southern Hemisphere is measured counter-clockwise from the right of the diagram as if you were looking at the Earth down from the North Pole. *b*, Variance reduction plotted against direction of the symmetry axis of a uniformly anisotropic inner core with cylindrical symmetry. Black represents a variance reduction of $>70\%$; contour interval is 10% . White is less than 40% .



anisotropy in a layer extending from about 100 to 300 km depth in the inner core, and weak anisotropy outside, may go a long way towards explaining all the observations. Recall that the rays analysed here turn ~ 100 – 300 km beneath the inner-core boundary, whereas those analysed by Shearer *et al.*¹¹ turn either above or below this region. An alternative explanation, suggested by the observation made here that P'_{DF} rays with $\cos^2 \xi > 0.6$ have anomalously low amplitudes, is that these phases are systematically difficult to observe above ambient noise. This may lead to unreliable travel-time picks in the ISC Bulletin for this sampling direction. An analysis of differential times and amplitude ratios measured from digital waveforms of P'_{CD} minus P'_{DF} (ref. 17) in the distance range 130 – 140° , and P'_{AB} minus P'_{DF} in the distance range 146 – 180° , may help resolve the depth dependence of anisotropy.

Implications

My primary result is that 92% of the P'_{BC} minus P'_{DF} differential travel-time residuals are small (r.m.s. scatter of 0.4 s), whereas the mean of the remaining 8% of the times is 2.9 s, seven standard deviations away from the 'normal' observations. The geographical sampling of anomalous and normal paths overlaps (Fig. 4*a*), but all of the anomalous observations (early, small-amplitude DF) have inner-core ray paths nearly parallel to the Earth's spin axis. The observations can be explained (69% variance reduction of 264 independent observations) using a two-parameter model of uniform, cylindrically symmetric anisotropy. These data only sample the outer 300 km of the inner core, and the anomalous paths sample mainly latitudes within 45° of the Equator but are widely, although not uniformly, distributed in longitude. There is therefore a strong case for P-wave anisotropy at low latitudes in the outermost inner core which is 3.5% fast parallel to the Earth's spin axis. From differential times alone, we cannot distinguish between inner-core anisotropy advancing the times of DF or anisotropy in a boundary layer at the base of the outer core delaying the times of BC. Inner core anisotropy is, however, consistent with the absolute time observations of P'_{DF} (refs 10, 11), and with the sign of anomalous mode splitting, whereas the outer-core boundary layer hypothesis would produce mode splitting of opposite sign to the observations. In addition to anisotropy, there seems to be large-scale heterogeneity which is not dominantly axisymmetric (Fig. 4*a* and *b*) and is small in amplitude, giving rise to travel-time anomalies typically less than 0.5 s. This could be caused by structure in the inner core or a boundary layer at the base of the outer core.

The symmetry axis of models of cylindrically symmetric uniform anisotropy that best fits the observations is 5° from the Earth's spin axis, and the locus of all symmetry axes that provide reasonable fits (say $>50\%$ variance reduction) includes the spin axis, but only 6% of the range of possible directions (Fig. 5*b*).

These observations constrain the symmetry axis to be so near to the rotation axis that their coincidence is unlikely to be due to random chance.

The inner core seems to be made primarily of the hexagonally close-packed ϵ phase of iron¹⁸, whose compressional wave-speeds are anisotropic. The magnitude of anisotropy increases with pressure, but decreases with temperature, so its value extrapolated to inner-core conditions is highly uncertain. Two mechanisms for aligning the iron crystals are well organized large-scale convective flow and preferred alignment as the inner core freezes. If the temperature of the inner core is near the solidus, it may be convectively unstable¹⁹. The lowest mode of convection of a homogeneous sphere heated from within is characterized by flow straight through the centre with return flow along the entire surface²⁰. If the axis of symmetry of this flow is the spin axis, one obtains nearly uniform simple shear at low latitudes which could align the iron crystals in a well organized manner²¹. This mechanism would not produce the mode of uniform anisotropy proposed here but the anisotropy would be roughly uniform in the region sampled by the anomalous paths, and produce a model that is consistent with the differential times. Alternatively, anisotropy could have been frozen in during inner-core formation and aligned relative to the spin axis (maximum moment of inertia). For the inner core to be aligned now this principal axis of inertia would have had to remain unchanged over the growth time of the outer 300 km of the inner core, perhaps as long as 10^9 years.

The amplitude ratios BC/DF of the anomalous phases are three times larger than the ratio measured for 'normal' paths at the same epicentral distances. This ratio commonly exceeds 15. The small-amplitude DF phases make travel-time picks of the anomalous phases rare and prone to error because DF is seldom above the noise. Because of its consistency and large magnitude, the amplitude observations may place strong constraints on core structure, which will be analysed elsewhere (T.J.M. and K.C.C., manuscript in preparation). Possible explanations are that high velocities in the inner core advance the times and cause anomalous geometrical spreading, reducing amplitudes. Alternatively, the transmission and reflection coefficients at the inner-core boundary may be strongly affected by the anisotropic properties of the inner core. Further analysis of waveforms of P'_{DF} along ray directions nearly parallel to the Earth's spin axis is needed to improve the resolution of inner-core structure.

Received 12 August 1991; accepted 3 March 1992.

1. Masters, G. & Gilbert, F. *Geophys. Res. Lett.* **8**, 569–571 (1981).
2. Ritzwoller, M., Masters, G. & Gilbert, F. *J. geophys. Res.* **91**, 10203–10228 (1986).
3. Ritzwoller, M., Masters, G. & Gilbert, F. *J. geophys. Res.* **93**, 6369–6396 (1988).
4. Woodhouse, J. H., Giardini, D. & Li, X.-D. *Geophys. Res. Lett.* **13**, 1549–1552 (1986).
5. Giardini, D., Li, X.-D. & Woodhouse, J. H. *Nature* **325**, 405–411 (1987).
6. Giardini, D., Li, X.-D. & Woodhouse, J. H. *J. geophys. Res.* **93**, 13716–13742 (1988).
7. Li, C.-D., Giardini, D. & Woodhouse, J. H. *J. geophys. Res.* **96**, 551–577 (1991).

8. Widmer, R. thesis, Univ. of California, San Diego (1991).
9. Poupinet, G., Pillet, R. & Souriau, A. *Nature* **305**, 204–206 (1983).
10. Morelli, A., Dziewonski, A. M. & Woodhouse, J. H. *Geophys. Res. Lett.* **13**, 1545–1548 (1986).
11. Shearer, P. M., Toy, K. M. & Orcutt, J. A. *Nature* **333**, 228–232 (1988).
12. Shearer, P. M. & Toy, K. M. *J. geophys. Res.* **96**, 2233–2247 (1991).
13. Dziewonski, A. M. & Anderson, D. L. *Phys. Earth planet. Inter.* **25**, 297–356 (1981).
14. Dziewonski, A. M. & Gilbert, F. *Geophys. J. R. astr. Soc.* **44**, 7–17 (1976).
15. Kennett, B. L. N. & Engdahl, E. R. *Geophys. J. Int.* **105**, 429–465 (1991).
16. Boyd, T. M. & Creager, K. C. *J. geophys. Res.* **96**, 2267–2291 (1991).
17. Cormier, V. F. & Choy, G. L. *Geophys. Res. Lett.* **13**, 1553–1556 (1986).
18. Brown, J. M. & McQueen, R. G. *J. geophys. Res.* **91**, 7485–7494 (1986).
19. Jeanloz, R. & Wenk, H. R. *Geophys. Res. Lett.* **15**, 72–75 (1988).
20. Chandrasekhar, S., *Hydrodynamic and Hydromagnetic Stability* (Clarendon, Oxford, 1961).
21. Wenk, H. R., Takeshita, T., Jeanloz, R. & Johnson, G. C. *Geophys. Res. Lett.* **15**, 76–79 (1988).

ACKNOWLEDGEMENTS. I thank P. M. Shearer for preprints, code to read GDSN data from CD-ROMs and a review and R. T. Merrill and J. M. Brown for discussions. The digital seismic data came primarily from the GDSN operated by the US Geological Survey. This work was supported by the NSF.

Lymphoproliferation disorder in mice explained by defects in Fas antigen that mediates apoptosis

Rie Watanabe-Fukunaga*, Camillynn I. Brannan†, Neal G. Copeland†, Nancy A. Jenkins† & Shigekazu Nagata*‡

* Osaka Bioscience Institute, 6-2-4 Furuedai, Suita-shi, Osaka 565, Japan

† Mammalian Genetics Laboratory, ABL-Basic Research Program, NCI-Frederick Cancer Research and Development Center, Frederick, Maryland 21702, USA

Fas antigen is a cell-surface protein that mediates apoptosis. It is expressed in various tissues including the thymus and has structural homology with a number of cell-surface receptors, including tumour necrosis factor receptor and nerve growth factor receptor. Mice carrying the lymphoproliferation (*lpr*) mutation have defects in the Fas antigen gene. The *lpr* mice develop lymphadenopathy and suffer from a systemic lupus erythematosus-like autoimmune disease, indicating an important role for Fas antigen in the negative selection of autoreactive T cells in the thymus.

T-CELL precursors arise in the bone marrow and then mature in the thymus after interaction with the thymic microenvironment¹. During maturation, T cells recognizing self-antigens are destroyed by a process called apoptosis, whereas others are positively selected^{2,3}. The lymphoproliferation mutation (*lpr*) seems to interfere with T-cell maturation. The *lpr* is autosomal recessive and the phenotype includes formation of multiple autoantibodies and accumulation of large numbers of non-malignant CD4⁺CD8[−] T lymphocytes in lymph nodes and the spleen⁴. Two independent spontaneous *lpr* mutations have been identified, *lpr* and *lpr*^{cg}. The original *lpr* mutation occurred during the derivation of the MRL/MpJ strain⁵ and has been transferred onto a number of other inbred strain backgrounds, including C3H and C57BL⁶. The other mutation is *lpr*^{cg} in the CBA/K1Jms mouse strain⁷. The clinical syndrome of *lpr* and *lpr*^{cg} mice is characterized by hypergammaglobulinaemia, anti-DNA antibodies, rheumatoid factor, and circulating immune complexes as well as arthritis and glomerulonephritis, which resembles human systemic lupus erythematosus (SLE)⁴. Studies of *lpr* mice suggest that there is a defect in negative selection of self-reactive T lymphocytes in the thymus^{8,9}. Excessive numbers of self-reactive T lymphocytes released into peripheral organs seem to be responsible for the autoimmune disease in *lpr* mice.

Here we provide evidence that *lpr* encodes the structural gene

for the mouse Fas antigen¹⁰. A mouse monoclonal antibody has been prepared which has a cytolytic activity on human cells expressing the Fas antigen¹¹. Cloning of Fas antigen complementary DNA from human¹² and mouse cells¹⁰ indicates that the Fas antigen is a protein containing a single transmembrane domain with a calculated *M_r* of 35,000 (35 K) (refs 10, 12). Fas antigen from both species shows structural homology with a number of cell-surface receptors, including tumour necrosis factor (TNF) receptors and the low-affinity nerve growth factor receptor^{10,12}. Northern analysis indicates the Fas antigen messenger RNA is expressed in a limited number of tissues, including the thymus, liver, ovary and heart¹⁰. When human Fas antigen is expressed in mouse cell lines, it can induce Fas antigen antibody-triggered cell death¹². Characterization of the process of cell death indicates that Fas antigen mediates apoptosis¹². Expression of Fas antigen in the thymus and its role in apoptosis provides an explanation for the phenotypes of *lpr* mice. These studies further suggest a role for the Fas antigen in the negative selection of autoreactive T cells in the thymus.

Expression of Fas mRNA in *lpr* mice

The mouse Fas antigen gene has been assigned to chromosome 19 by interspecific backcross analysis¹⁰. When our linkage map of mouse chromosome 19 was aligned with a composite linkage map (GBASE, Jackson Laboratory) we found that the Fas antigen locus mapped near *lpr*¹³.

To determine whether the *lpr* mutation affects Fas antigen mRNA expression, northern analysis of wild-type and mutant *lpr* tissues was done using a mouse Fas antigen cDNA probe. Total cellular RNAs were isolated from two different strains carrying the *lpr* mutation (MRL *lpr/lpr* and C3H *lpr/lpr*) and their parental wild-type controls (MRL +/+ and C3H +/+). In agreement with previous observations¹⁰, a 2.1 kilobase (kb) Fas antigen mRNA is detected in the liver and thymus of mice wild type at *lpr* (Fig. 1a). Almost no Fas mRNA is observed in homozygous *lpr* mice. Reprobing the same blot with human elongation factor-1α (EF-1α) cDNA¹⁴ reveals a band of 2.2 kb in all RNA preparations (Fig. 1b). To confirm the absence of Fas mRNA in *lpr* mice, single-stranded cDNAs were synthesized from thymus RNA of wild-type and *lpr* mice and amplified by the polymerase chain reaction (PCR), using Fas antigen oligonucleotide primers. PCR amplification of RNA from wild-type mice gave a band of the expected size (420 bp), whereas no such band was amplified from *lpr* mice (Fig. 1c).

‡ To whom correspondence should be addressed.

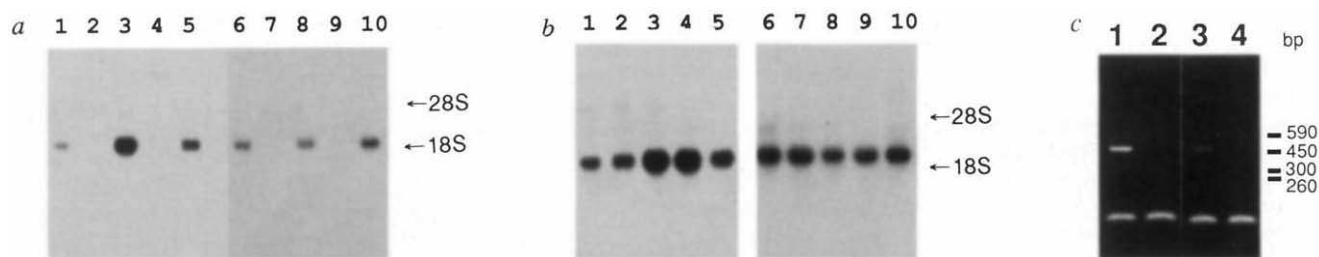


FIG. 1 Little expression of Fas antigen mRNA in mice homozygous for the *lpr* mutation. Northern blot analysis of total cellular RNA from various mouse strains hybridized with mouse Fas antigen cDNA¹⁰ (a) or human EF-1 α cDNA¹⁴ (b). Total cellular RNAs (9 μ g) are from the liver (lanes 1, 2, 6, 7) and thymus (lanes 3, 4, 8, 9) of MRL/MpJ +/+ (lanes 1, 3), MRL/MpJ *lpr/lpr* (lanes 2, 4), C3H/HeJ +/+ (lanes 6, 8), and C3H/HeJ *lpr/lpr* (lanes 7, 9) mice. Total cellular RNA (4.5 μ g) from mouse L929 cells was also analysed (lanes 5, 10). Positions of 18S and 28S ribosomal RNAs are indicated. c, The Fas antigen mRNA detected by reverse PCR. Single-stranded cDNA was synthesized using total cellular RNAs from the thymus of MRL/MpJ +/+ (lane 1), MRL/MpJ *lpr/lpr* (lane 2), C3H/HeJ +/+ (lane 3) and C3H/HeJ *lpr/lpr* (lane 4) mice, and Fas antigen cDNA was amplified by PCR. Size markers (base pairs) are shown on the right.

METHODS. Total cellular RNAs, prepared by the guanidine isothiocyanate/acid phenol method²⁹, were denatured at 65 °C for 5 min in 2.2 M formaldehyde

and 50% deionized formamide, and electrophoresed through 1.5% agarose gels containing 2.2 M formaldehyde³⁰. RNA was transferred to a nitrocellulose filter and hybridized with [³²P] labelled probe DNA. The probe DNAs used are a 1.5-kb *Eco*RI fragment containing mouse Fas antigen cDNA¹⁰, or a 1.8-kb *Bam*HI fragment containing human EF-1 α cDNA¹⁴. The reverse PCR was done as described³¹. Random hexamer-primed single-stranded cDNA was synthesized in a final volume of 50 μ l with 80 units avian myeloblastosis reverse transcriptase and 2 μ g total RNA. An aliquot (5 μ l) of the reaction mixture was diluted with 50 μ l of the PCR buffer³², and the PCR reaction was done using 2.5 units *Thermus aquaticus* DNA polymerase (*Taq* polymerase). The conditions for the PCR were 1.5 min at 95 °C, 1.5 min at 60 °C, and 3 min at 72 °C, 30 cycles. The primers spanned nucleotides 45–64 and nucleotides 448–467 in the coding sequence of mouse Fas antigen cDNA¹⁹.

Rearrangement of the Fas antigen gene

To determine whether the absence of Fas antigen mRNA in *lpr* mice is associated with a rearrangement of the Fas antigen gene, Southern analysis of genomic DNAs from mutant *lpr* and wild-type mice was done using a full-length Fas antigen cDNA probe. Several bands were observed in *Eco*RI-, *Bam*HI- and *Pst*I-digested genomic DNA from wild-type MRL and C3H mice (Fig. 2a, b). Except for an extra 9-kb band observed in *Bam*HI-digested DNA from C3H mice (Fig. 2b, lane 1), all bands were identical in size between the two strains. The genomic locus for the wild-type mouse Fas antigen gene has been isolated (R.W.F. and S.N., unpublished observations). A comparison of the restriction map of the cloned genomic locus with DNA fragments observed in Southern analysis indicates that the Fas antigen gene is present in a single copy per haploid mouse genome and that some bands detected by Southern analysis represent closely migrating doublets.

Southern analysis of DNA from MRL *lpr/lpr* mice showed extra bands of 8 kb or 9 kb in *Bam*HI- or *Eco*RI-digested DNAs, respectively (Fig. 2a, lanes 3, 6). The 5-kb *Pst*I fragment in wild-type DNA (Fig. 2a, lane 7) was replaced by a slightly larger fragment (Fig. 2a, lane 9). When DNAs from wild-type and mutant strains were mixed before restriction enzyme digestion, bands expected for both wild-type and mutant strains were

observed (Fig. 2a, lanes 2, 5, 8), indicating that the extra bands seen in *lpr* DNA are not due to artefacts generated during restriction enzyme digestion. Southern analysis of DNA from C3H *lpr/lpr* mice gave bands identical to those in MRL *lpr/lpr* mice (Fig. 2b). These results indicate that the structural gene for the Fas antigen is rearranged in *lpr* mice.

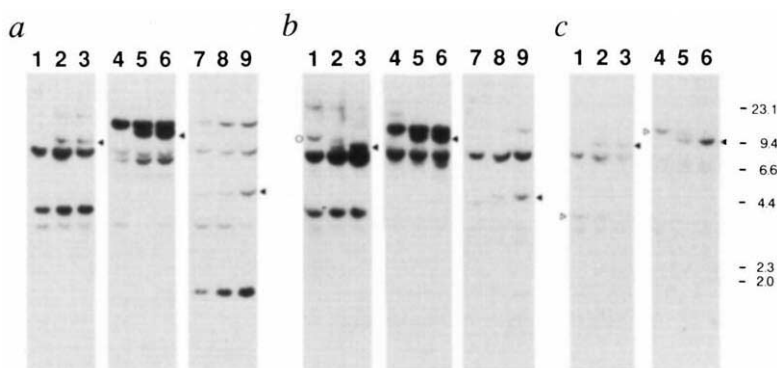
To determine which region of the Fas antigen gene was rearranged, mouse Fas antigen cDNA was divided into four parts and each part was used as a probe for Southern analysis of wild-type and *lpr* DNAs. Rearrangement of the Fas antigen gene was detected only with the probe that contains nucleotides 115–580 of mouse Fas antigen cDNA¹⁰. The 4-kb *Bam*HI and 10-kb *Eco*RI DNA fragments seen in wild-type DNA were replaced by 8-kb and 9-kb fragments, respectively, in *lpr* DNA (Fig. 2c). Preliminary analysis of Fas antigen genomic DNA clones suggests that the 4-kb *Bam*HI fragment detected in wild-type DNA contains exon 3 and that the rearrangement seen in *lpr* mice has occurred in intron 2 (R.W.F. and S.N., unpublished observations).

A point mutation in Fas antigen

To confirm that Fas antigen is the structural gene for *lpr*, a second *lpr* mutation, *lpr*^{cs}⁷, was analysed. Northern analysis of RNAs from liver and thymus of CBA *lpr*^{cs}/*lpr*^{cs} mice indicates

FIG. 2 Southern analysis of the Fas antigen gene in various strains. High *M*_r DNA was prepared from the spleen of a, c, MRL/MPJ and b, C3H/HeJ mice. Genomic DNA (10 μ g per lane) from wild-type (lanes 1, 4, 7) and *lpr/lpr* mutant mice (lanes 3, 6, 9), and their 1:1 mixture (lanes 2, 5, 8) was digested with *Bam*HI (lanes 1–3), *Eco*RI (lanes 4–6) or *Pst*I (lanes 7–9). Southern analysis was done using a full-length mouse Fas antigen cDNA¹⁰ (a, b) or coding sequences between nucleotides 115–580 (c) as probes. Filled arrowheads, DNA fragments rearranged in *lpr* mice; open arrowheads, the corresponding DNA fragments in wild-type mice; Open circle, the polymorphic DNA fragment observed in *Bam*HI-digested C3H DNA.

METHODS. High *M*_r chromosomal DNA, prepared from the spleen³⁰, was digested with various restriction enzymes, electrophoresed on a 0.8% agarose gel, and transferred to a nylon membrane under alkaline conditions as described³⁰. Probe DNA fragments were prepared by PCR using mouse Fas antigen cDNA¹⁰ as a template, and labelled



with ³²P by the random primer labelling method (Boehringer). Hybridization was carried out under high stringency³⁰. Sizes of marker DNA are shown in kb on the right.

that this mutant mouse expresses Fas antigen mRNA to the same extent as wild-type CBA $+/+$ mice (Fig. 3a). After reverse transcription of CBA $+/+$ RNA, the coding sequence of the Fas antigen cDNA was amplified by PCR and inserted into pBluescript. The PCR product has a sequence identical to that of mouse Fas antigen cDNA isolated from the mouse BAM3 cell line¹⁰ (derived from BALB/c mice), except at nucleotide positions 162, 163, 283, 379 and 505. These differences probably represent polymorphisms between BALB/c and CBA mice. The sequence of three independent Fas antigen cDNA clones, obtained with RNA from CBA lpr^{cg}/lpr^{cg} mice, all show a transition of T to A at nucleotide 786, in addition to the polymorphic mutations mentioned above. This mutation causes the replacement of an asparagine for an isoleucine in the cytoplasmic region of the Fas antigen (Fig. 3b), which is highly conserved between the Fas antigen and TNF receptor type I.

To establish whether this mutation abolishes the ability of the Fas antigen to transduce the apoptotic signal, and to overcome the lack of antibodies against the mouse Fas antigen, two chimaeric Fas antigen molecules were constructed from human and mouse cDNAs. Both chimaeras carry the extracellular and transmembrane domains of the human Fas antigen cDNA. The cytoplasmic domains are derived from Fas antigen cDNA of wild-type or lpr^{cg} mice, respectively. The chimaeric cDNAs were expressed in mouse L929 cells, and the susceptibility of the transformants to the Fas antibody-induced apoptosis was assayed¹².

Transformants expressing the chimaera with the wild-type Fas antigen cytoplasmic tail were effectively killed by the Fas antibody (Fig. 4), indicating that the cytoplasmic region of the wild-type mouse Fas antigen can transduce the apoptotic signal into cells. Cells expressing the chimaera with the mutated mouse Fas antigen from lpr^{cg} mice were resistant to the cytolytic activity of the anti-human Fas antibody. These results indicate that the Fas antigen encoded by lpr^{cg} mutant mice is unable to transduce the apoptotic signal into cells, and that lpr encodes Fas antigen.

Discussion

We have shown that lpr mice carry defects in the Fas antigen which mediates apoptosis¹². Bone marrow transplantation studies indicate that the lpr mutation causes an intrinsic T-

lymphocyte abnormality responsible for lymphadenopathy and autoantibody production¹⁵. The Fas antigen has been expressed in the thymus¹⁰ and on activated or transformed T cells¹². Thus it is possible that precursor T cells reacting with self-antigens express the Fas antigen during their thymic development and are killed by interaction with stroma cells in the thymus. No functional Fas antigen is expressed in lpr mice. Therefore, autoreactive lpr T cells could escape thymic selection, resulting in the lymphadenopathy and autoimmune disease in lpr mice^{4,15}.

Mice carrying the generalized lymphoproliferative disease mutation (*gld*) show a clinical syndrome which is indistinguishable from that of lpr mice^{6,16}. Like lpr , *gld* is autosomal recessive but *gld* maps to mouse chromosome 1¹⁷. Bone marrow transplantation¹⁸ indicates that these two mutations are expressed by different cell compartments and affect an interacting pair of molecules, suggesting that the structural gene for *gld* may encode the ligand for Fas antigen. The molecule affected by the *gld* mutation is thought to be expressed in bone marrow-derived cells¹⁸. That the thymic stromal cells responsible for T-cell depletion are derived from the bone marrow^{1,19,20} supports this.

Transgenic mice expressing a rearranged T-cell receptor for the H-Y male antigen and carrying the lpr mutation have been constructed^{9,21}. Although male lpr mice carrying the transgene show an increased number of autoreactive T-cells in the spleen and an enhanced production of autoantibodies⁹ relative to female mice, clonal deletion of autoreactive T-cells was still observed. These results indicate that in addition to the Fas antigen, other molecules have a role in the clonal deletion of autoreactive T-cells, although it is not possible to rule out the weak expression of the Fas antigen in lpr mice. The involvement of defective B-cells in the autoimmune disease of lpr mice has also been suggested^{22,23}. Because Fas antigen can be expressed in activated B lymphocytes¹¹, B-cell producing autoantibodies could survive longer in lpr mice than normal mice. Such an extension of the functional lifespan of B lymphocytes could cause autoimmune disease in a similar way to that proposed for transgenic mice carrying an unregulated *bcl-2* gene²⁴.

Transfer of bone marrow stem cells from lpr mice into irradiated wild-type mice causes a severe graft-versus-host-like disease (GVHD)²⁵. But this is not seen when bone marrow stem cells from lpr^{cg} mice are transferred into irradiated wild-type

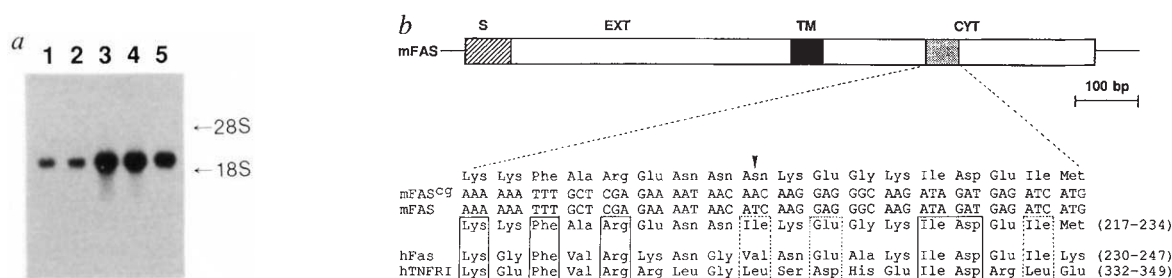


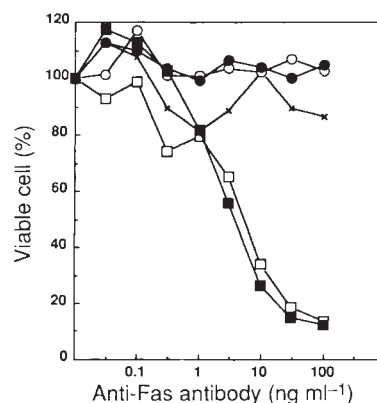
FIG. 3 Structure of the Fas antigen gene in mice homozygous for the lpr^{cg} mutation. a, Northern analysis of Fas antigen mRNA in lpr^{cg} mice. CBA/K1Jms $+/+$ and CBA/K1Jms lpr^{cg}/lpr^{cg} mice total cellular RNA (9 μ g) prepared from the liver (lanes 1, 2) or thymus (lanes 3, 4) of wild-type (lanes 1, 3) or mutant (lanes 2, 4) mice was analysed using mouse Fas antigen cDNA as probe. Total RNA from L929 cells (lane 5) was also analysed. Positions of 18S and 28S rRNA are indicated. b, Point mutation in the cytoplasmic region of the Fas antigen gene of lpr^{cg}/lpr^{cg} mice. The coding region of the Fas antigen gene expressed in the thymus and liver of wild-type and mutant mice was amplified by PCR, and the nucleotide sequence determined after insertion into pBluescript. Upper panel, Predicted structure of the mouse Fas antigen protein. S, signal sequence; EXT, extracellular region; TM, transmembrane region; CYT, cytoplasmic region. Lower panel, Nucleotide sequence and deduced amino-acid sequence of the Fas antigen cDNA of lpr^{cg} (mFas^{cg}) and wild-type (mFas) mice at the site of the mutation. Amino-acid sequences of the corresponding region of the human Fas antigen (hFas)¹² and human TNF type I receptor (hTNFRI)³² are included. Sets of three identical amino-acid

residues at one aligned position are boxed by solid lines, sets of three residues regarded as favoured substitutions are boxed by dotted lines. The arrowhead indicates the position of the mutation in the Fas antigen of lpr^{cg} mice.

METHODS. Total RNA (2 μ g) from the liver or thymus of CBA $+/+$ or CBA lpr^{cg}/lpr^{cg} mice was used as a template for single-stranded cDNA synthesis. PCR was carried out as described³¹. The oligonucleotides used as primers are GGAATTCGCTGTTTCCCTTGCTGCA (5' primer) and GGTCGACCAGGAGTTGCCAATGTCAAT (3' primer), which contains the sequence in the 5' or 3' non-coding region of the mouse Fas antigen cDNA¹⁰, respectively. PCR conditions were 1.5 min at 95 °C, 1.5 min at 60 °C, and 3 min at 72 °C, 50 cycles. The amplified DNA fragment was digested with *Eco*RI and *Sal*I, and inserted into pBluescript KS(+). The sequencing reaction was done by the dideoxynucleotide chain-termination method using T7-DNA polymerase (Pharmacia) and [α -³²S]dATP α S (Amersham). As primers for the sequencing reaction, a set of 20-base oligonucleotides specific for the coding region of the mouse Fas antigen gene was used.

FIG. 4 Inability of the Fas antigen encoded by *lpr*^{cr} mice to mediate apoptosis. Mouse L929 cell transformed by and expressing a chimaeric human-mouse Fas antigen containing the wild-type mouse Fas antigen (□, ■) or the mutated mouse Fas antigen (○, ●) cytoplasmic region were incubated at 37 °C for 6 h with various concentrations (0.03–100 ng ml⁻¹) of anti-human Fas antibody in the presence of 0.5 μg ml⁻¹ actinomycin D. After incubation, viable cells were stained with crystal violet, and dye uptake was quantified by measuring absorbance at 540 nm using an automated Micro-ELISA autoreader. The parental L929 cells (x) were also treated with anti-human Fas antibody in the presence of actinomycin D. Assays were done in duplicate for two independent cell lines transformed by and expressing each chimaeric Fas antigen, and the average values for each transformant were plotted.

METHODS. A hybrid cDNA between mouse and human Fas antigen cDNAs was constructed by the recombinant PCR method³³. The primary PCR was done with two sets of primers. A part of the human Fas antigen cDNA was amplified with a forward primer (primer A) containing human sequence from 696–715¹², and a 41-base hybrid primer (primer B) containing the sequence complementary to mouse Fas antigen cDNA from 586–608¹⁰, and to human Fas antigen cDNA from 725–743¹². A part of mouse Fas antigen cDNA was amplified using the wild-type mouse Fas antigen cDNA or the mutated cDNA from *lpr*^{cr} mice as template. The oligonucleotide complementary to primer B, and a 20-base oligonucleotide (primer C) complementary to the sequence from 984–1,003 of the mouse Fas antigen cDNA¹⁰ were used as primers. The conditions for the PCR were 1 min at 94 °C, 2 min at 55 °C and 2 min at 72 °C, 15 cycles. Products were isolated by agarose gel electrophoresis, and treated with T4 DNA polymerase to flush the ends. The two DNA fragments obtained by the primary PCRs were then mixed at 1:1, and the secondary PCR was done for 20 cycles as above using primers A and C. The product was digested with *Bam*HI and *Kpn*II, and the 330-bp DNA



fragment was isolated. This fragment was ligated with the 700-bp *Xba*I-*Bam*HI fragment containing the 5' part of human Fas antigen cDNA¹² and the 660-bp *Kpn*II-*Xba*I fragment containing the 3' part of mouse Fas antigen cDNA¹⁰, and introduced into mammalian expression vector pEF-BOS³⁴. All constructions were confirmed by DNA sequencing. The expression plasmid for the human-mouse chimaeric Fas antigen was introduced into mouse L929 cells¹². After 9 days, individual G-418-resistant colonies were isolated, and clones transformed by and expressing the chimaeric Fas antigen were identified by FACS analysis using mouse anti-human Fas monoclonal antibody.

mice²⁶. These two observations can be explained as follows. Fas antigen mRNA is not expressed in *lpr* mice. When *lpr* stem cells introduced in wild-type mice become mature T-cells, they recognize the Fas antigen expressed in the host cells as a foreign antigen or 'non-self' and induce the GVHD syndrome. Mice carrying the *lpr*^{cr} allele express the Fas antigen, although in a non-functional form. Therefore, mature T-cells derived from stem cells of *lpr*^{cr} mice can recognize host cells expressing the normal Fas antigen as 'self'.

Our results explain the phenotype observed in double heterozygous *lpr*+/+, *gld*+/+, and *lpr*^{cr}+/+, *gld*+/+ mice⁷. Although the original *lpr* mutation fails to complement *gld* in double heterozygotes, *lpr*^{cr} can complement *gld*; *lpr*^{cr}+/+, *gld*+/+ double heterozygotes develop a disease similar to that observed in *lpr*/*lpr* or *gld*/*gld* mice, although it is less pronounced. These phenotypic differences can be explained by the nature of the Fas antigen lesions found in *lpr* and *lpr*^{cr} mice. In *lpr* mice, Fas antigen is not expressed and can thus not bind ligand. In *lpr*^{cr} mice, a non-functional receptor is expressed that can still presumably bind ligand. In *lpr*^{cr}+/+, *gld*+/+ double heterozygotes,

the non-functional receptor could compete for ligand, further lowering the effective concentration of ligand below that normally found in *gld*+/+ mice. This reduced concentration of ligand could produce a phenotype similar to that seen in *lpr*/*lpr* or *gld*/*gld* mice.

Fas antigen is expressed in the thymus and in the liver, heart and ovary¹⁰. Abnormalities of these tissues have not been described in *lpr* mice, although CD4⁺ CD8⁺ T-lymphocytes which accumulate in *lpr* mice seem to proliferate in the liver²⁷. Because the liver can support extrathymic development of T-lymphocytes if the thymus is inactive or absent²⁷, it is possible that cells expressing Fas antigen in the liver are precursor cells for T-lymphocytes.

We have provided evidence that the *lpr* locus encodes Fas antigen. This finding should help to elucidate the process regulating normal T-cell maturation in the thymus, and the mechanism underlying formation of autoimmune disease in mice. Finally, human patients displaying a similar phenotype as *lpr* mice have been identified²⁸ and it will be interesting to examine the structure of the Fas antigen gene in these patients. □

Received 7 January; accepted 19 February 1992.

- van Ewijk, W. A. *Rev. Immun.* **9**, 591–615 (1991).
- von Boehmer, H. A. *Rev. Immun.* **6**, 309–326 (1988).
- Schwartz, R. H. *Cell* **57**, 1073–1081 (1989).
- Cohen, P. L. & Eisenberg, R. A. *Rev. Immun.* **9**, 243–269 (1991).
- Murphy, E. D. & Roths, J. B. in *Genetic Control of Autoimmune Disease* (eds Rose, N. R., Bigazzi, P. E. & Warner, N. L.) 207–221 (Elsevier/North-Holland, New York, 1978).
- Izui, S. *et al. J. Immun.* **133**, 227–233 (1984).
- Matsuzawa, A. *et al. J. exp. Med.* **171**, 519–531 (1990).
- Matsumoto, K. *et al. J. exp. Med.* **173**, 127–136 (1991).
- Zhou, T. *et al. J. Immun.* **147**, 466–474 (1991).
- Watanabe-Fukunaga, R. *et al. J. Immun.* (in press).
- Yonehara, S., Ishii, A. & Yonehara, M. *J. exp. Med.* **169**, 1747–1756 (1989).
- Itoh, N. *et al. Cell* **66**, 233–243 (1991).
- Watanabe, T. *et al. Biochem. Genet.* **29**, 325–336 (1991).
- Uetsuki, T., Naito, A., Nagata, S. & Kaziro, Y. *J. biol. Chem.* **264**, 5791–5798 (1989).
- Katagiri, T., Cohen, P. L. & Eisenberg, R. A. *J. exp. Med.* **167**, 741–751 (1988).
- Roths, J. B., Murphy, E. D. & Eicher, E. M. *J. exp. Med.* **159**, 1–20 (1984).
- Seldin, M. F. *et al. J. exp. Med.* **167**, 688–693 (1988).
- Allen, R. D., Marshall, J. D., Roths, J. B. & Sidman, C. L. *J. exp. Med.* **172**, 1367–1375 (1990).
- Marrack, P. *et al. Cell* **53**, 627–634 (1988).
- Jenkinson, E. J., Jaftay, P., Kingston, R. & Owen, J. J. T. *Transplantation* **39**, 331–333 (1985).

- Mountz, J. D., Zhou, T., Eldridge, J., Berry, K. & Bluthmann, H. *J. exp. Med.* **172**, 1805–1817 (1990).
- Perkins, D. L., Glaser, R. M., Mahon, C. A., Michaelson, J. & Marshak-Rothstein, A. *J. Immun.* **145**, 549–555 (1990).
- Sobel, E. S. *et al. J. exp. Med.* **173**, 1441–1449 (1991).
- Strasser, A. *et al. Proc. natn. Acad. Sci. U.S.A.* **88**, 8661–8665 (1991).
- Theofilopoulos, A. N. *et al. J. exp. Med.* **162**, 1–18 (1985).
- Kimura, M. *et al. Eur. J. Immun.* **21**, 63–69 (1991).
- Ohteki, T., Seki, S., Abo, T. & Kumagai, K. *J. exp. Med.* **172**, 7–12 (1990).
- Frizzera, G., Kaneko, Y. & Sakurai, M. *Leukemia* **3**, 1–5 (1989).
- Chomczynski, P. & Sacchi, N. *Analyt. Biochem.* **162**, 156–159 (1987).
- Sambrook, J., Fritsch, E. F. & Maniatis, T. in *Molecular Cloning: A Laboratory Manual*, 2nd edn (Cold Spring Harbor Laboratory Press, New York, 1989).
- Fukunaga, R., Seto, Y., Mizushima, S. & Nagata, S. *Proc. natn. Acad. Sci. U.S.A.* **87**, 8702–8706 (1990).
- Schall, T. J. *et al. Cell* **61**, 361–370 (1990).
- Higuchi, R. in *PCR Protocols, A Guide to Methods and Application* (eds Innis, M. A., Gelfand, D. H., Shinsky, J. J. & White, T. J.) 177–183 (Academic, San Diego, 1990).
- Mizushima, S. & Nagata, S. *Nucleic Acids Res.* **18**, 5322 (1990).

ACKNOWLEDGMENTS. We thank T. Katagiri, A. Matsuzawa and M. Yasumizu for *lpr* mice; S. Yonehara for anti-Fas antibody; N. Itoh and R. Fukunaga for discussion and comments; M. Ikeda for secretarial assistance. This work was supported by the Ministry of Education, Science and Culture of Japan, and the National Cancer Institute.

Molecular gas content of the primaeval galaxy IRAS 10214+4724

P. M. Solomon*, S. J. E. Radford† & D. Downes†

* Astronomy Program, State University of New York, Stony Brook, New York 11794, USA

† Institut de Radio Astronomie Millimétrique, 38406 St Martin d'Hères, France

THE faint IRAS source 10214+4724, identified¹ with a distant galaxy at redshift $z=2.286$, is one of the most intrinsically luminous objects in the Universe, with $L \approx 10^{14} L_{\odot}$ (where $L_{\odot}$ is the solar luminosity). The remarkable detection of emission from neutral CO (ref. 2) leads, by comparison with similar detections in nearby galaxies, to an estimate of $2\text{--}6 \times 10^{11} M_{\odot}$ for the mass of neutral molecular hydrogen in this galaxy. This is 30–90 times less than the estimate given in ref. 2, but still comparable to the total mass (gas, dust and stars) of a large spiral galaxy. This gas mass is consistent with the dynamical mass of the galaxy inferred from the CO emission line-width. The CO line luminosity is twenty times larger than is seen in nearby ($z < 0.3$) ultra-luminous ($L > 3 \times 10^{11} M_{\odot}$) galaxies^{3–5}. Although it is extremely gas-rich, with a higher CO luminosity than any other galaxy, the infrared to CO luminosity ratio of 10214+4724 is twice that of most infrared-luminous galaxies, and ~ 30 times higher than that of normal spirals. Its infrared colours and high infrared/CO luminosity ratio indicate that 10214+4724 is similar to nearby ultra-luminous galaxies⁵ that are known to be merging. This galaxy is a primaeval molecular galaxy with the mass of a large spiral, but with most of the mass in molecular gas rather than stars.

Monochromatic luminosity $L(\nu_{\text{rest}})$, observed flux $S(\nu_{\text{obs}})$ and luminosity distance D_L are related by⁶ $\nu_{\text{rest}} L(\nu_{\text{rest}}) = 4\pi D_L^2 \nu_{\text{obs}} S(\nu_{\text{obs}})$, where the redshifted frequency $\nu_{\text{obs}} = \nu_{\text{rest}}/(1+z)$. From the measured², velocity-integrated CO(3 $\rightarrow$ 2) line flux, $S\Delta v = 21 \text{ Jy km s}^{-1}$, we find that the line luminosity $L_{\text{CO}}(3\text{--}2) = \int L(\nu_{\text{rest}}) d\nu_{\text{rest}} = 4\pi D_L^2 S\Delta v \nu_{\text{obs}} = 1.7 \times 10^8 h^{-2} L_{\odot}$ ($h = H_0/100 \text{ km s}^{-1} \text{ Mpc}^{-1}$, $q_0 = 0.5$; Table 1). (Brown and Vanden Bout² miscalculated a luminosity $(1+z)^2 = 11$ times too large). Examination of the published² spectrum indicates that the CO(3 $\rightarrow$ 2) flux is uncertain by $\sim 40\%$, so all parameters derived from the line flux are similarly uncertain. Because $S \propto T_b \nu^2$, then $L \propto \nu^3$ if the brightness temperature T_b is constant, which is characteristic of thermalized, opaque lines. The CO(1 $\rightarrow$ 0) line should therefore have a flux of 2.3 Jy km s^{-1} , nine times less than the CO(3 $\rightarrow$ 2) line, and a luminosity $L_{\text{CO}}(1\text{--}0) = 6 \times 10^6 h^{-2} L_{\odot}$.

Often CO line luminosities are expressed as the product of the velocity-integrated brightness temperature, $T_b \Delta v$, of the source, and its area, $\Omega_s D_A^2$, where Ω_s is the solid angle subtended by the source and the angular size distance⁶ $D_A = D_L/(1+z)^2$. For a source smaller than the beam, the observed integrated line intensity, I_{CO} , measures the beam-diluted brightness temperature, which also decreases with redshift, $T_b \Delta v \Omega_s = I_{\text{CO}} \Omega_b (1+z)$, where Ω_b is the solid angle of the telescope beam. Then the line luminosity is

$$L'_{\text{CO}} = T_b \Delta v \Omega_s D_A^2 = \Omega_b D_L^2 I_{\text{CO}} / (1+z)^3$$

In these terms, 10214+4724 has a CO(3 $\rightarrow$ 2) luminosity $L'_{\text{CO}}(3\text{--}2) = 1.3 \times 10^{11} h^{-2} \text{ K km s}^{-1} \text{ pc}^2$ and $L'_{\text{CO}}(1\text{--}0) = L'_{\text{CO}}(3\text{--}2)$ if the brightness temperatures are equal. For comparison, 10214+4724 has a CO luminosity 20 times higher than the most gas-rich ultra-luminous infrared galaxies known³, 20087–0308 and 15030+4835, and 30 times more than the most CO-luminous normal spiral galaxies⁷, objects such as NGC3147 or NGC7771. We note that the observed integrated line intensity

for a given telescope beam area is

$$I_{\text{CO}} = L'_{\text{CO}}(1+z)^3 / (\Omega_b D_L^2)$$

The factor of $(1+z)^3$ implies that detection of thermal emission lines at high redshift is easier than had been previously thought, particularly if transitions from a higher rotational level at higher frequency are used, keeping the beam size roughly constant.

If the excitation parameters are known, the H_2 mass can be estimated^{4,8} from the line luminosity:

$$M(H_2)/L'_{\text{CO}} = \alpha = (4\mu/3\pi G)^{1/2} (n(H_2)^{1/2}/T_b)$$

where $(4\mu/3\pi G)^{1/2} = 2.1 M_{\odot} (\text{K km s}^{-1} \text{ pc}^2)^{-1} \text{ cm}^{3/2} \text{ K}$ and $n(H_2)$ is the H_2 density. This linear relation has been empirically verified^{9,10} for CO(1 $\rightarrow$ 0) in the Galaxy, where typically $n(H_2) \approx 500 \text{ cm}^{-3}$ and $T_b \approx 10 \text{ K}$, so $\alpha = 4.6 M_{\odot} (\text{K km s}^{-1} \text{ pc}^2)^{-1}$. With this ratio we find $M(H_2) \approx 6 \times 10^{11} h^{-2} M_{\odot}$ for 10214+4724. Using this same α , Brown and Vanden Bout² overestimated $M(H_2)$ because they used redshifted CO(3 $\rightarrow$ 2) flux in a formula for non-redshifted CO(1 $\rightarrow$ 0), thus obtaining a mass $(1+z) \times (\nu_{32}/\nu_{10})^2 = 30$ times too large. The correct expression is

$$M(H_2) = \alpha (c^2/2k) D_L^2 S\Delta v / [\nu_{\text{rest}}^2 (1+z)]$$

or for a CO transition from $J \rightarrow J-1$,

$$M(H_2) = 1.1 \times 10^4 M_{\odot} \left[\frac{\alpha}{4.5} \right] \left[\frac{D_L}{\text{Mpc}} \right]^2 \times \left[\frac{S\Delta v}{\text{Jy km s}^{-1}} \right] \left[\frac{1}{J^2(1+z)} \right]$$

where α is in units of $M_{\odot} (\text{K km s}^{-1} \text{ pc}^2)^{-1}$. The true $M(H_2)$ may be even smaller because the excitation conditions of molecular gas in an extremely luminous galaxy at $z = 2.286$ will be significantly different from those in the Milky Way. In particular, the brightness temperature may be much higher as the cosmic background radiation at that epoch was hotter, $(1+z) \times 2.75 \text{ K} = 9 \text{ K}$, the infrared measurements¹ indicate very warm dust and the mere presence of significant emission in the CO(3 $\rightarrow$ 2) line indicates a substantial population in the $J = 3$ state, 33 K above the ground level. If, for example, the rest-frame brightness temperature $T_b = 50 \text{ K}$ and $n(H_2) = 1,000 \text{ cm}^{-3}$, then α and the mass will be smaller by a factor of three, giving $M(H_2) = 1.7 \times 10^{11} h^{-2} M_{\odot}$. A more accurate mass estimate is difficult, but it is clear that 10214+4724 has a CO luminosity at least 20 times that of gas-rich galaxies in the local Universe and that its H_2 mass is roughly equal to the total mass of a large spiral galaxy like the Milky Way. This suggests 10214+4724 is a primarily gaseous, molecular galaxy with a normal mass, but seen at an early evolutionary stage. Although galaxies in the present epoch have 90% stars and 10% gas, these proportions may be reversed in 10214+4724.

Three clues to the nature of 10214+4724 are the ratio of gas and dynamical masses, the infrared-to-CO luminosity ratio and the mid-to-far infrared luminosity ratio. If the observed optical radius¹, $R = 2'' = 8 h^{-1} \text{ kpc}$, indicates the extent of the molecular gas, then the CO line-width², $\Delta v = 250 \text{ km s}^{-1}$, implies the total dynamical mass $M_{\text{dyn}} = R\Delta v^2/G \sin^2 i = 2.3 \times 10^{11} h^{-1} M_{\odot}$ for a rotating disk at an inclination of $i = 45^\circ$ to the plane of the sky. This implies

$$\frac{M(H_2)}{M_{\text{dyn}}} = 0.75 \left[\frac{0.5}{\sin^2 i} \right] \left[\frac{8 \text{ kpc}}{R} \right] \left[\frac{\alpha}{1.3} \right]$$

Other nearby ultra-luminous galaxies have similarly large $M(H_2)/M_{\text{dyn}}$ ratios¹¹, but the spatial extent of the gas in those systems is only 0.6–2 kpc. If the CO source in 10214+4724 radiates as a black body, the minimum size $R_{\text{min}} = (L_{\text{CO}}/\pi T_b \Delta v)^{1/2} = 1,800 h^{-1} \text{ pc}$ for $T_b = 50 \text{ K}$. This minimum size implies $M_{\text{dyn}} > 5 \times 10^{10} h^{-1} M_{\odot}$ for $i = 45^\circ$, or $M(H_2)/M_{\text{dyn}} < 3$. To reconcile the masses in this case requires

TABLE 1 Properties of IRAS 10214 + 4724 at $z = 2.286$

D_L	$8.8 h^{-1}$	Gpc	
S_{100}	550 ± 150	mJy	ref. 1
S_{60}	190 ± 40	mJy	ref. 1
S_{25}	75 ± 31	mJy	ref. 1
$L(6-36 \mu\text{m})$	$7.5 \times 10^{13} h^{-2}$	$L_\odot$	ref. 1
$\nu_{\text{obs}}(\text{CO}(3 \rightarrow 2))$	105.233	GHz	ref. 2
$S_{\text{CO}}(3-2)\Delta\nu$	21	Jy km s $^{-1}$	ref. 2
$\Delta\nu$ (FWHM)	250	km s $^{-1}$	ref. 2
$L_{\text{CO}}(3-2)$	$17 \times 10^7 h^{-2}$	$L_\odot$	
$L_{\text{CO}}(1-0)$	$6 \times 10^6 h^{-2}$	$L_\odot$	if $T_b(3-2) = T_b(1-0)$
$L'_{\text{CO}}(3-2)$	$1.3 \times 10^{11} h^{-2}$	K km s $^{-1}$ pc 2	
$L'_{\text{CO}}(1-0)$	$1.3 \times 10^{11} h^{-2}$	K km s $^{-1}$ pc 2	if $T_b(3-2) = T_b(1-0)$
$M(\text{H}_2)$	$1.7 \times 10^{11} h^{-2}$	$M_\odot$	if $\alpha = 1.3$
M_{dyn}	$2.3 \times 10^{11} h^{-1}$	$M_\odot$	if $i = 45^\circ$, $R \times 8$ kpc
$L_{100 \mu\text{m}}$	$1.3 \times 10^{13} h^{-2}$	$L_\odot$	inferred from CO luminosity

The luminosity distance $D_L = ch_0^{-1} q_0^{-2} [z q_0 (q_0 - 1) [(2q_0 z + 1)^{0.5} - 1]]$ where $h = H_0/100$ km s $^{-2}$ Mpc $^{-1}$ and we take $q_0 = 0.5$. The line luminosity, $L = 4\pi D_L^2 S \Delta\nu \nu_{\text{obs}}/c$, where $S \Delta\nu$ is the velocity-integrated line flux and ν_{obs} is the redshifted line frequency, can also be expressed as $L' = \Omega_b D_L^2 I_{\text{CO}} / (1+z)^3$, where I_{CO} is the velocity-integrated line intensity and Ω_b is the beam solid angle. These two expressions for the luminosity are related by $L = (8\pi \nu_{\text{rest}}^3 / c^3) L'$.

a smaller inclination, $i < 25^\circ$, or a higher brightness temperature, $T_b > 50$ K.

The infrared-to-CO luminosity ratio is a measure of how efficiently the molecular gas fuels star formation^{7,12}. For 10214 + 4724, the rest-frame far-infrared luminosity has not been measured, but the mid-infrared luminosity¹, $L(6-36 \mu\text{m}) = 7.5 \times 10^{13} h^{-2} L_\odot$, provides a lower limit of $L_{\text{IR}}/L'_{\text{CO}} > 580 L_\odot$ (K km s $^{-1}$ pc 2) $^{-1}$. Ultra-luminous infrared galaxies, such as Arp 220 or the Seyfert 1 quasar Mrk 231, typically have^{3,13} $L_{\text{IR}}/L'_{\text{CO}} = 100-400 L_\odot$ (K km s $^{-1}$ pc 2) $^{-1}$, where $L_{\text{IR}} = L(8-1,000 \mu\text{m})$. In our sample, partially described previously³, of 30 interacting ultra-luminous infrared galaxies with $z < 0.3$ and $3 \times 10^{11} < L < 3 \times 10^{12} h^{-2} L_\odot$, we find $(L_{\text{IR}}/L'_{\text{CO}}) = 190 \pm 50$, and a sample of 11 warm ultra-luminous galaxies^{5,13} has $(L_{\text{IR}}/L'_{\text{CO}}) = 250$. Isolated normal spiral galaxies typically have⁷ $(L_{\text{IR}}/L'_{\text{CO}}) = 32$, almost an order of magnitude less than ultra-luminous galaxies. Thus the $L_{\text{IR}}/L'_{\text{CO}}$ ratio of 10214 + 4724 is 30 times the ratio typical for isolated spiral galaxies, and two to three times the average ratio for ultra-luminous galaxies. One previously measured¹⁴ infrared galaxy, 08572 + 3915, has $L_{\text{IR}}/L'_{\text{CO}}$ as high as 10214 + 4724. In its ratio of infrared luminosity to H_2 mass, 10214 + 4724 resembles other ultra-luminous infrared galaxies, but not normal spirals. It seems closest to scaled-up versions of the warmest (in IRAS 25 μm , 60 μm and 100 μm colours) luminous infrared galaxies⁵.

Our sample of 30 ultra-luminous galaxies shows a constant ratio, to within a factor of two, of CO(1 $\rightarrow$ 0) line flux to observed 100 μm flux, $S_{\text{CO}}(1-0)\Delta\nu/S_{100 \mu\text{m}} = 3.5$ km s $^{-1}$, corresponding to a CO-to-100 μm luminosity ratio $L_{\text{CO}}(1-0)/\nu L_\nu(100 \mu\text{m}) = 4.5 \times 10^{-7}$. Only ultra-luminous galaxies show such a tight correlation between these fluxes. Applying our empirical result to 10214 + 4724, we infer a flux of 580 mJy at 329 μm (rest frame 100 μm), which implies $\nu L_\nu(100 \mu\text{m}) = 1.3 \times 10^{13} L_\odot$, 20% of the mid-infrared luminosity. This inferred flux falls close to model B of ref. 1 for dust emission from a starburst in a massive spiral galaxy¹, which has $4 \times 10^8 h^{-2} M_\odot$ of dust. If we assume, for comparison, a temperature of 50 K for dust radiating at 100 μm

and a standard dust emissivity, we find a dust mass of $10^9 h^{-2} M_\odot$.

The large mid-to-far infrared luminosity ratio in 10214 + 4724 shows the galaxy is unusually warm. From the measured¹ 100- μm flux and the 330- μm estimated above, we infer that the rest-frame 30 $\mu\text{m}/100 \mu\text{m}$ ratio is 0.95. Our sample of ultra-luminous galaxies is much cooler, with 25 $\mu\text{m}/100 \mu\text{m}$ ratios ranging from 0.05 to 0.029. Galaxies in the warm sample⁵ do have higher ratios, ~ 0.5 , but only one¹⁴, Mrk 463, is as hot as 10214 + 4724. The ultra-luminous Seyfert 2 cD galaxy⁵ 09104 + 4109, to which 10214 + 4724 was initially compared¹, also has a very warm spectrum.

Like the quasar Mrk 1014, which has warm infrared colours and a narrow CO line (190 km s $^{-1}$ full width at half maximum), but unlike Arp 220, which has very cool infrared colours and extraordinarily broad CO lines (800 km s $^{-1}$ full width at half maximum), 10214 + 4724 has a rather narrow CO line-width. This suggests we are seeing the inner disks of 10214 + 4724 and Mrk 1014 more nearly face-on than that of Arp 220. If the mid-infrared radiation is emitted by a hot central source surrounded by a disk of cooler material, the apparent mid-infrared luminosity will be very sensitive to the inclination of the galaxy because the flux will decrease exponentially with the opacity along the line of sight, whereas the CO and far infrared, tracing the cooler envelope, are relatively insensitive to inclination.

What is 10214 + 4724? Apart from its impressive bolometric luminosity, its outstanding characteristic is its mass of molecular gas, which we have shown is equal to the total mass of a large spiral galaxy. Its CO luminosity is what we would expect of a galaxy composed primarily of molecular clouds. It may be a rotating disk galaxy at an inclination of 20–45° in an early stage of evolution. Although a previous generation of stars has synthesized the heavy elements making up the dust grains where H_2 forms, most of the mass remains as gas not yet formed into stars. Whereas 10214 + 4724 has an infrared luminosity/ H_2 mass ratio twice that of some other, nearer ultra-luminous galaxies, normal galaxies have ratios 35 times smaller. Its infrared and CO properties resemble most closely those of warm ultra-luminous galaxies⁵, most of which are interacting or merging systems. It may be a merger or simply a galaxy in the process of disk formation.

Can star formation power 10214 + 4724, or is an active galactic nucleus a more likely explanation? A possible test of the star formation theory is to observe HCN, which traces gas at least 10 times denser than CO. Recently we have shown¹⁶ that ultra-luminous infrared galaxies have extremely high HCN luminosities. If star formation in dense clouds powers 10214 + 4724, those observations suggest it should have a HCN luminosity which is 0.1% of its infrared luminosity, or almost equal to its CO luminosity. Formation of high-mass stars consumes molecular gas at a rate¹⁷ $\dot{M}/L_{\text{IR}} = 10^{-10} M_\odot \text{ yr}^{-1} L_\odot^{-1}$, so the starburst lifetime, $M(\text{H}_2)/\dot{M} = 3 \times 10^7$ yr, would be a few per cent of the age of the Universe at $z = 2.3$. The large gas mass, which is comparable to the total mass of a galaxy in the present Universe, the absence of broad emission lines¹ and a ratio of infrared luminosity to molecular mass comparable to other starburst galaxies all indicate that star formation is the power source, transforming a primaeval molecular galaxy into stars. $\square$

Received 16 December 1991; accepted 14 February 1992.

- Rowan-Robinson, M. et al. *Nature* **351**, 719–721 (1991).
- Brown, R. L. & Vanden Bout, P. A. *Astr. J.* **102**, 1956–1959 (1991).
- Downes, D., Solomon, P. M. & Radford, S. J. E. in *Dynamics of Galaxies and their Molecular Cloud Distributions* (eds Combes, F. & Casoli, F.) 295–298 (Kluwer, Dordrecht, 1991).
- Radford, S. J. E., Solomon, P. M. & Downes, D. *Astrophys. J.* **369**, L15–L18 (1991).
- Sanders, D. B., Soifer, B. T., Elias, J. H., Neugebauer, G. & Matthews, K. *Astrophys. J.* **328**, L35–L39 (1988).
- Weinberg, S. *Gravitation and Cosmology* (Wiley, New York, 1972).
- Solomon, P. M. & Sage, L. J. *Astrophys. J.* **334**, 613–625 (1988).
- Solomon, P. M., Radford, S. J. E. & Downes, D. *Astrophys. J.* **348**, L53–L56 (1990).

- Solomon, P. M., Rivolo, A. R., Barrett, J. & Yahil, A. *Astrophys. J.* **319**, 730–741 (1987).
- Scoville, N. Z. et al. *Astrophys. J. Suppl.* **63**, 821–915 (1987).
- Scoville, N. Z., Sargent, A. I., Sanders, D. B. & Soifer, B. T. *Astrophys. J.* **366**, L5–L9 (1991).
- Young, J. et al. *Astrophys. J.* **304**, 443–458 (1986).
- Sanders, D. B. in *Dynamics of Galaxies and their Molecular Cloud Distributions* (eds Combes, F. & Casoli, F.) 417–423 (Kluwer, Dordrecht, 1991).
- Sanders, D. B. et al. *Astr. Astrophys.* **213**, L5–L8 (1989).
- Kleinmann, S. G. et al. *Astrophys. J.* **328**, 161–169 (1988).
- Solomon, P. M., Downes, D. & Radford, S. J. E. *Astrophys. J.* **387**, L55–L61 (1992).
- Scoville, N. Z. & Young, J. S. *Astrophys. J.* **265**, 148–165 (1983).

The origin of planets orbiting millisecond pulsars

Marco Tavani*† & Leigh Brookshaw*

* Institute of Geophysics and Planetary Physics, LLNL,
University of California, California 94550, USA

† Department of Astronomy, University of California at Berkeley,
Berkeley, California 94720, USA

At least two Earth-sized planets have been discovered around the 6-ms pulsar PSR1257+12 (ref. 1), which, like millisecond pulsars in general, has probably been spun up by accretion of material from a companion star. In addition, two 'star-vaporizing' millisecond pulsars (SVPs), 1957+20 (ref. 2) and 1744–24A (refs 3, 4), show evidence of mass outflows from their low-mass companions, which are thought to be vaporized by pulsar radiation. Building on this, we suggest a model for the formation of planets around millisecond pulsars such as 1257+12, which no longer have stellar companions. We present detailed hydrodynamical models which suggest that planet formation can occur either in a low-mass X-ray binary progenitor to a progenitor of an SVP when the neutron star is accreting material driven off its companion by X-ray irradiation (refs 5, 6), or after a pulsar has formed and is vaporizing its companion^{5,7–9}. In both cases a circum-binary disk is created in which planets can form on a timescale of 10^5 – 10^6 years¹⁰ (which is short compared with the binary evolution timescales of the parent systems) and the planets can survive a second phase in which the companion star moves towards the pulsar and is completely vaporized⁵.

The model proposed here for planet formation around millisecond pulsars is a natural consequence of irradiation-driven evolution of low-mass X-ray binaries (LMXBs) and SVPs⁵ as qualitatively suggested in ref. 1 in the case of PSR1257+12 and in ref. 9 in the case of PSR1957+20.

Let us discuss the SVP case first. The discovery of 'star-vaporizing pulsars' such as PSR1957+20 and PSR1744–24A confirmed the relevance of star irradiation and consequent mass loss on the binary evolution of millisecond pulsars⁵. The pulsar wind interacts with the mass outflow and forms a termination shock where electron-positron pairs may be efficiently accelerated and emit synchrotron radiation¹¹. The radiation spectrum depends on the details of the shock acceleration mechanism and on the composition of the pulsar wind (see, for example, ref. 12). The effect of star irradiation is to produce a large temperature ($T \approx 10^6$ – 10^7 K) mass outflow^{5,13} which is reprocessed by a shock occurring where the pulsar radiation pressure balances the gas pressure. Complex hydrodynamical processes contribute to give the final pressure and velocity configuration of the material which then moves out from the binary. Detailed calculations of the mass outflows of PSR1957+20 and PSR1744–24A (refs 7, 8) and the eclipse properties of these two systems show that the mass outflow configuration can have different shapes and that the pulsar can be completely enshrouded¹⁴ by the evaporated material, as occasionally happens for PSR1744–24A (ref. 3).

We discuss here results of hydrodynamic models of SVP mass outflow which may lead to the formation of planetary systems. We used a version of the smoothed particle hydrodynamic (SPH) method¹⁵ which incorporates the pulsar radiation pressure in addition to gravitational and pressure forces⁷. We followed the computational method of ref. 7, and the material is assumed to be optically thick to pulsar electromagnetic radiation except for the leading mass particles which are the closest to the pulsar. We have considered several models characterized by different masses and orbital distances of the companion star as well as different mass loss rates $|\dot{m}|$, thermal and kinematic properties of the mass outflow, and pulsar luminosities. We present here a selected sample of our results relevant to the mechanism of planetary formation.

Results of a two-dimensional calculation of the outflow can be applied to the material that resides near the orbital plane. A comprehensive three-dimensional hydrodynamic calculation of the outflow at large radial distances is not currently feasible. Our preliminary three-dimensional SPH models up to distances comparable with the orbital radius suggest, however, that a small fraction $f \leq 0.1$ of the outflowing material may be recaptured or reside in the orbital plane for a variety of initial conditions. Recent observations¹⁶ with the Very Large Array suggest that this is the case for PSR1957+20.

In our calculations we assumed wind velocities of the order of the escape velocity from the companion star of mass $m \leq 0.1 M_\odot$, Mach numbers of order unity, orbital distances of order 10^{11} cm, and a wind efficiency of conversion of irradiating energy into outflow kinetic energy η of less than a few per cent⁵. We find that for relatively large irradiating energy fluxes, for example $|\dot{m}| \approx 10^{15}$ – 10^{16} g s^{–1}, and pulsar luminosities $L_p \approx 10^{35}$ erg s^{–1} (typical of systems such as PSR1957+20), the shock at the inner binary tends to produce a velocity that forces the evaporated material to escape from the binary to infinity with a typical two-wing shaped spiral (see Fig. 1). Material in the 'inner' wing interacts with the 'outer' wing, and additional shocks may occur at large distances.

We find, however, that for smaller irradiating fluxes (for example, $L_p \leq 10^{34}$ erg s^{–1} and $|\dot{m}| \approx 10^{14}$ g s^{–1}) a substantial fraction of the mass outflow remains bound at large distances even after having been 'reprocessed' by the pulsar termination shock. The pulsar may in this case be totally or partially enshrouded and the outflowing material is shielded from the direct action of the pulsar wind. Figure 2 shows the results of a two-dimensional hydrodynamic model characterized by a Roche-lobe-filling companion of mass $m = 0.1 M_\odot$, pulsar luminosity $L_p \approx 2 \times 10^{33}$ erg s^{–1}, initial gas temperature at injection on the irradiated half of the companion's surface $T = 5 \times 10^6$ K, Mach number $M = 1$, and $\eta \approx 2 \times 10^{-2}$. Figure 2a shows the details of the mass outflow near the binary, and Fig. 2b gives a large-scale view of the same mass outflow with the inclusion of a ballistic trajectory representing the behaviour of non-interacting mass particles with velocities and initial directions

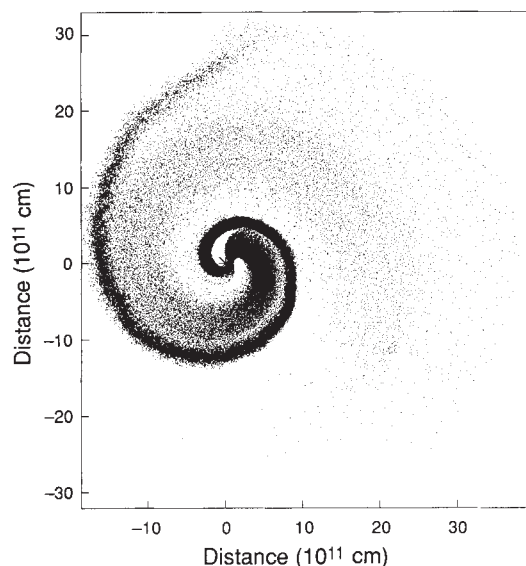


FIG. 1 SVP with mass outflow escaping to infinity, 1.6 orbital periods after starting the simulation. The pulsar companion is filling its Roche radius, with $m = 0.1 M_\odot$. The initial temperature of the outflow is $T_i = 10^7$ K, the mass-loss rate is $\dot{m} = 10^{15}$ g s^{–1}, the initial velocity of the gas is $M = 1.5$, the period of the binary orbit is 5 h and the luminosity of the pulsar is $L = 8 \times 10^{33}$ erg s^{–1}. The pulsar is at the point (0,0) and the orbital motion is counterclockwise. The units of length are in 10^{11} cm. The plot is in the corotating frame.

obtained from the hydrodynamic calculation. We note that the outflowing material has in this case negative total energy and that the maximum radius reached is ~ 60 times the original orbital radius.

In a realistic case, the mass particles cannot spiral inwards because of the interaction with the mass outflow, and therefore the outflowing material in the orbital plane can accumulate in a circum-binary disk for a relatively long time. For a disk mass $m_d \approx 30$ Earth masses (from the properties of PSR1257+12), we obtain the estimate of the mass-loss rate from the companion as $|\dot{m}|_1 \approx (10^{14} \text{ g s}^{-1})/(\tau_8 \tau_e)$, where $\tau_e = (10^8 \text{ yr})\tau_8$ is the SVP evolutionary timescale (see below). The orbital angular momentum of the companion in the original orbit is $J \approx (8.48 \times 10^{50} \text{ erg s})m_{-1}\sqrt{a_{11}}$, where $m_{-1} = m/(10^{-1} M_\odot)$ and $a_{11} = a/(10^{11} \text{ cm})$ with a the orbital distance. On the other hand, the total angular momentum of the two planets orbiting PSR1257+12 is $J_p \approx 2 \times 10^{48} \text{ erg s}$, smaller than the original binary angular momentum by a factor $\sim km_{-1}\sqrt{a_{11}}$ with $k \approx 2$ for an ideal solar composition of the circum-binary material, and with $k \approx 20$ for material originating from a helium companion. Therefore, the fraction of the initial angular momentum that needs to be stored in a circum-binary disk depends on the chemical composition of the companion. Heavy elements are more likely to be gravitationally bound and form planets, whereas the remaining H, He, and C, N, O elements may be evaporated away from the binary system. The rest of the angular momentum is redistributed by inward binary motion and outward mass outflow.

Angular momentum transport in the mass outflow may occur because of turbulent viscosity induced by hydrodynamic shocks and, possibly, by magnetic instabilities¹⁷. The timescale for viscous transport of angular momentum is $\tau_\nu \approx r^2/\nu$, where r is the radius and ν the viscosity, and it is expected to be of order $\sim 10^6$ – 10^7 yr for the densities, temperatures and distances involved in outer disks (see also refs 18, 19). The angular momentum transported outwards influences the binary orbit, which is therefore expected to shrink as in the case of PSR1957+20 (ref. 20). The expanding gas cools because of adiabatic losses and radiative losses. The upper limit to the temperature at a distance $a_{11} \approx 100$ obtained from adiabatic losses is $T \leq (2 \times 10^3 \text{ K}) T_{i,6}$, where $T_{i,6}$ is the initial temperature in units of 10^6 K , and additional shocks in the outflow will decrease the temperature even further because of radiative losses. The pulsar may be completely enshrouded and therefore 'hidden'^{14,21,22}, and the outflowing gas is exposed only to the high-energy radiation (X-rays and γ -rays) produced near the termination shock of the pulsar wind^{12,14}. The evaporated gas at large distance from the binary is expected to be only marginally affected by high-energy radiation as detailed calculations suggest for PSR1744–24A (ref. 8). In this case, a gravitationally bound circum-binary disk is likely to be formed. The heating from the central source does not influence the large-scale hydrodynamics of the circum-binary disk in SVP systems with hidden pulsars, and the outflowing material eventually reaches a radius $a_{11} \approx 100$ at which additional molecular cooling and dust formation occurs. At this distance, grains can coalesce and coagulate rapidly^{23,24}, and planetesimal formation is expected on a timescale $\tau_p \approx 10^6$ yr (ref. 19). Planet formation therefore naturally tends to occur at a distance from the central source of order 1 AU, although one cannot exclude the possibility that planets form at larger distances.

A complete treatment of viscous and resonant torques acting on the SVP binary evolution is beyond our scope here. It is plausible, however, that because of the angular momentum loss, the binary progressively contracts on a timescale $\tau_c \approx 10^7$ – 10^8 yr as in the case of PSR1957+20 (ref. 20). This is smaller than the typical spin-down timescale of millisecond pulsars τ_s ($\tau_s \approx 10^9$ yr for PSR1257+12; ref. 1) and comparable with τ_ν . In this case, the companion star is exposed to a larger and larger radiation flux which enhances both the mass loss from its surface and the pressure forces at the inner shock. The effect of the termination

shock during this second evolutionary phase, in which the companion moves closer to the central pulsar, is to increase the velocity of the outflowing gas. The latest stages of the star-vaporization process may be characterized by relatively large mass-loss rates ($\sim 10^{16} \text{ g s}^{-1}$) and outflow velocities larger than the binary escape velocity. No material is expected to be gravitationally bound in a circum-binary disk in this evolutionary phase. The original companion star can be completely evaporated in a timescale less than τ_s (ref. 5), and only the previously coagulated planets or planetesimals can witness the process of pulsar-driven star evaporation.

Alternatively, planet formation can occur in radiation-driven low-mass X-ray binaries (RD-LMXBs) with evolutionary timescales $\tau_e \approx 10^7$ yr (ref. 25) whose mass-loss rates are enhanced by X-ray irradiation^{5,6,13,21} and that turn at a later stage into SVPs⁵. Again a circum-binary disk can form at large distance if a substantial fraction of the mass outflow is gravitationally bound. The mass-loss rates involved in this case are $|\dot{m}| \approx$

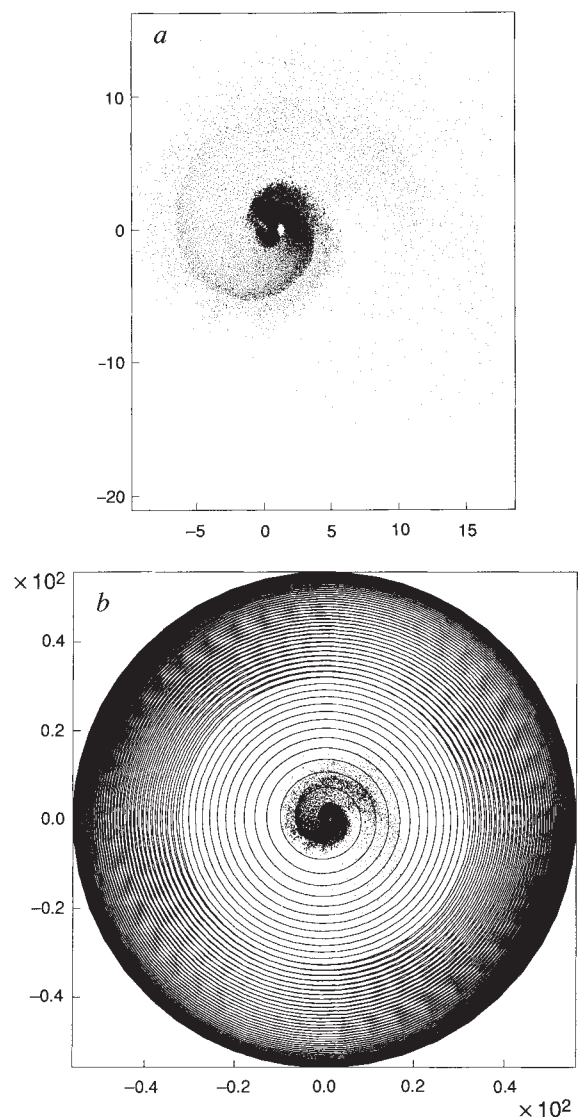


FIG. 2 SVP with a gravitationally bound outflow, 1.6 orbital periods after starting the simulation (compare with Fig. 1). The initial temperature of the outflow is $T_i = 5 \times 10^6 \text{ K}$, the mass-loss rate is $\dot{m} = 5 \times 10^{14} \text{ g s}^{-1}$, the initial velocity of the gas is $M = 1.0$, the period of the binary orbit is 5 h and the luminosity of the pulsar is $L = 2 \times 10^{33} \text{ erg s}^{-1}$. *a*, Outflowing material is trapped in the system (compare with Fig. 1). *b*, Ballistic trajectory superimposed over the hydrodynamic result of *a*. The initial conditions for the ballistic trajectory were calculated from a block of particles at position (0, -5).

$10^{18}\text{--}10^{19}\text{ g s}^{-1}$, and a small fraction of the mass loss of order $f \approx 0.1/k$ needs to be trapped in the system depending on the chemical composition of the outflowing material. We ran several two-dimensional and preliminary three-dimensional SPH models of LMXB outflows (whose results are to be reported elsewhere; manuscript in preparation) and obtained the necessary f for a variety of orbital parameters, $T = 10^7\text{ K}$ and $M \approx 1$. In this case, planet formation can proceed in a way similar to the mechanism suggested in ref. 19, with the crucial difference of having much more angular momentum available. Only at distances $\sim 10^{13}\text{ cm}$ from the central power source (now an X-ray source possibly screened by the accretion disk material) can the outflow material cool down to temperatures that allow grain and planetesimal formation. The irradiation-driven mass loss of RD-LMXBs is expected to be suddenly quenched after

$\sim 10^7\text{ yr}$ (refs 6, 21, 25), and such systems can result in binaries consisting of spun-up millisecond pulsars with low-mass companions⁵. The binary then turns into an SVP if the pulsar is sufficiently powerful and the companion is relatively close, and the evolution described above can be applied for the later stages of binary evolution. We note that the planets formed during a previous RD-LMXB phase can survive the second SVP phase.

An irradiation model of planetary formation may allow the existence of Earth-sized planets surrounding SVPs whose characteristics favour the gravitational binding of outflow material. Furthermore, the characteristics of the mass outflow make plausible the formation of planets at large distances from RD-LMXBs. Infrared emission might be detectable from circum-binary disks, and we urge systematic observations of SVPs and LMXBs. $\square$

Received 16 December 1991; accepted 14 February 1992.

1. Wolszczan, A. & Frail, D. A. *Nature* **355**, 145–147 (1992).
2. Fruchter, A. S., Stinebring, D. R. & Taylor, J. H. *Nature* **333**, 237 (1988).
3. Lyne, A. G. *et al.* *Nature* **347**, 650–652 (1990).
4. Nice, D. J., Thorsett, S. E., Taylor, J. H. & Fruchter, A. S. *Astrophys. J.* **361**, L61 (1990).
5. Ruderman, M., Shaham, J. & Tavani, M. *Astrophys. J.* **336**, 507 (1989).
6. Tavani, M. *Nature* **351**, 39 (1991).
7. Tavani, M. & Brookshaw, L. *Astrophys. J.* **381**, L21–L24 (1991).
8. Tavani, M. & Brookshaw, L. *Astrophys. J.* (submitted).
9. Banit, M. & Shaham, J. *Astrophys. J.* (in the press).
10. Safronov, V. S. & Ruzmaikina, T. V. in *Protostars & Planets* (eds Black, D. & Matthews, M.) 959–980 (University of Arizona Press, Tucson, 1985).
11. Kennel, C. F. & Coroniti, F. V. *Astrophys. J.* **283**, 694 (1984).
12. Arons, J. & Tavani, M. *Astrophys. J.* (submitted).
13. London, R. A., McCray, R. & Auer, L. H. *Astrophys. J.* **243**, 970 (1981).
14. Tavani, M. *Astrophys. J.* **379**, L69–L73 (1991).
15. Monaghan, J. J. *Computer Phys. Rep.* **3**, 71–124 (1985).

16. Fruchter, A. S. & Goss, M. *Astrophys. J.* **384**, L47–L51 (1992).
17. Balbus, S. & Hawley, J. *Astrophys. J.* **376**, 214 (1991).
18. Shu, F. H., Lubow, S. H. & Anderson, L. *Astrophys. J.* **229**, 223–241 (1979).
19. Lin, D. N. C., Woosley, S. E. & Bodenheimer, P. H. *Nature* **353**, 827 (1991).
20. Ryba, M. F. & Taylor, J. H. *Astrophys. J.* **380**, 557 (1991).
21. Tavani, M. in *X-Ray Binaries and the Formation of Binary and Millisecond Radio Pulsars* (eds van den Heuvel, E. P. J. & Rappaport, S.) (Kluwer, Dordrecht, in the press).
22. Tavani, M. in *Proc. 2nd GRO Science Workshop* (ed. Shrader, C.) (NASA, Greenbelt, in the press).
23. Goldreich, P. & Ward, W. R. *Astrophys. J.* **183**, 1051 (1973).
24. Safronov, V. S. *Evolution of Protoplanetary Cloud and the Formation of the Earth and Planets* (Nauka, Moscow, 1969); (Engl. transl. NASA TTF-677).
25. Tavani, M. *Astrophys. J.* **366**, L27–L31 (1991).

ACKNOWLEDGEMENTS. We thank A. Wolszczan for discussions. M.T. acknowledges discussions with M. Banit, M. Ruderman and J. Shaham. Work was done at IGPP-LNL under the auspices of the U.S. DOE and with NASA support.

Evolution and advection of solar mesogranulation

Richard Muller*, Hervé Auffret*, Thierry Roudier†, Jean Vigneau‡, George W. Simon‡, Zoe Frank§, Richard A. Shine§ & Alan M. Title§

* Observatoire du Pic du Midi, 65200 Bagnères de Bigorre, France

† Observatoire du Pic du Midi, 31400 Toulouse, France

‡ Phillips Laboratory (AFSC), National Solar Observatory, Sunspot, New Mexico 88349, USA

§ Lockheed Palo Alto Research Laboratory, Palo Alto, California 94304, USA

GRANULAR structure on the Sun's surface, with a typical scale of 1–2 Mm, has been known since 1800, and one hundred years ago, with the first observations by spectroheliograph^{1,2}, a mesh-like bright network was found with a characteristic scale of 30 Mm (40"). This pattern was found, thirty years ago, to be coincident with close-packed convective cells ('supergranulation') revealed by Doppler observations^{3–5} to be nestling inside the bright network. More recently^{6,7} an intermediate 'mesogranular' structure was found, with a characteristic scale of 3–10 Mm. We have obtained a three-hour sequence of observations at the Pic du Midi observatory which shows the evolution of mesogranules from appearance to disappearance with unprecedented clarity. We see that the supergranules, which are known to carry along (advect) the granules with their convective motion, also advect the mesogranules to their boundaries. This process controls the evolution and disappearance of mesogranules.

The origin of the bright network first seen by Hale¹ and Deslandres² was not explained until the observation of convective cells inside the mesh structure. It was then argued⁵ that convection sweeps magnetic flux to the boundaries of the cells and concentrates it there, enhancing local heating of the upper photosphere and chromosphere. A half-hour movie of the mesogranular structure made using the Solar Optical Universal

Photometer (SOUP) instrument on board the 1985 Spacelab 2 space shuttle flight⁷ first revealed the complexity of the surface flow patterns, which are evidence for subphotospheric convection. The SOUP movies showed that granules are advected by the larger-scale convective flows, and thus could be used as tracers of their motions. Horizontal flow maps produced by correlation tracking of the granules confirmed the existence of mesogranules and greatly improved our understanding of the intimate relation between supergranules and magnetic structures⁸. Unfortunately the SOUP movie lasted only 27.5 min (because of the day–night cycle of the space shuttle's equatorial near-Earth orbit), much less than the lifetime of supergranules and mesogranules, and its spatial resolution was limited by the relatively small size (30 cm) of the telescope aperture. Here we report recent observations taken at the Observatoire du Pic du Midi (French Pyrenees) with a 50-cm telescope, in a continuous 3-h sequence. The observations are nearly diffraction-limited.

Unfortunately, even the best ground-based observations are blurred and distorted by 'seeing' (turbulence in the Earth's atmosphere). To remove local image distortions, sophisticated algorithms have recently been developed to align and destretch sequential images⁷. In addition, the entire solar atmosphere oscillates with a range of periods centred around 5 min. These periods are sufficiently close to the lifetime of the granulation pattern ($\sim 10\text{ min}$) to complicate studies of convective flows. But because these oscillations have a well defined dispersion relation, and the phase velocity of the waves generally is higher than the sound speed, three-dimensional Fourier filtering in space–time⁷ is effective in removing them from a movie. After elimination of the seeing distortions and solar wave motions, local correlation tracking^{9,10} (a technique for measuring motions of the granules) on the time-series of images determines the horizontal velocity field. Because the majority of convective flow patterns visible at the surface are horizontal, measurements with the commonly used Doppler technique are ineffective, and local correlation tracking provides the only method for the study of such motions near the centre of the solar disk. This procedure yields the average velocity of the intensity pattern inside the

tracking window (whose full width at half maximum is usually 1–3 Mm, and thus represents the motion of 2–10 granules).

The data were collected on film with a camera that took bursts of frames every 20 s. They were obtained on 20 September 1988, in a $70'' \times 100''$ area of the Sun near disk centre. A $67'' \times 67''$ portion of the best frame from each burst is first digitized using a $1,024 \times 1,024$ digital CCD camera and then destretched, space-time filtered, and correlation tracked. Because of solar rotation, each of the digitized images covers a slightly different area; thus our analysis is restricted to an area of $58'' \times 48''$ common to all the frames. The extremely high quality of the images is confirmed by the appearance of numerous network bright points (NBPs) which are only visible when the resolution is $0.3''$ or better. The NBPs are relatively short-lived bright areas in the surface which are cospatial with magnetic flux tubes^{11,12}, and can thus be used to infer the locations of magnetic structures in an image of the surface. In Fig. 1a we show a $70'' \times 50''$ section of one of these granulation images and identify several of the prominent NBPs by three converging short lines. (Such photos contain many small transient bright features; NBPs are defined to be those with lifetimes exceeding 7 min.)

In Fig. 1b we show the flow vectors derived from correlation tracking a $58'' \times 48''$ portion of Fig. 1a. All the NBPs observed during the first 20 min of the observing sequence are superposed on the vector plot. The outer rectangle delineates the area included in Fig. 1a. The NBPs, which extend beyond the digit-

ized area of the image, and the flow vectors permit us to outline roughly the boundary of a supergranule, shown as a hand-drawn oval. The supergranule is ~ 30 Mm ($40''$) across. We identify mesogranules as the centres of diverging flow vectors (sources) in Fig. 1b. We have calculated the effect of the flow vectors on an initially uniform distribution of 'corks' (test particles used to simulate magnetic flux tubes), and confirm earlier results^{8,13} that all corks eventually end up in the supergranular network, many of them at or near the observed NBPs. Cork patterns calculated from velocity maps separated by 2–3 hours are almost unchanged on the scale of the supergranule during this time interval.

At the smaller scale of mesogranules, however, these patterns are very different after two hours, implying considerable changes in the mesogranulation. To study the characteristics of mesogranule evolution during the three hours in more detail, we show in the left column of Fig. 2 four flow maps averaged over 17 min at 54-min intervals. (The averaging time must be long enough to reduce measurement noise but short compared to mesogranule lifetimes.) The velocity vectors $\mathbf{v} = (v_x, v_y)$ are superposed on a grey-scale image of the divergence $\nabla = \partial v_x / \partial x + \partial v_y / \partial y$ computed from these flows. Bright areas (where the vectors point outward; that is, diverge) correspond to sources; in the dark areas (sinks), arrows point inward (converge). In the right column of Fig. 2 is a two-level contour map of the divergence. The levels displayed were chosen to

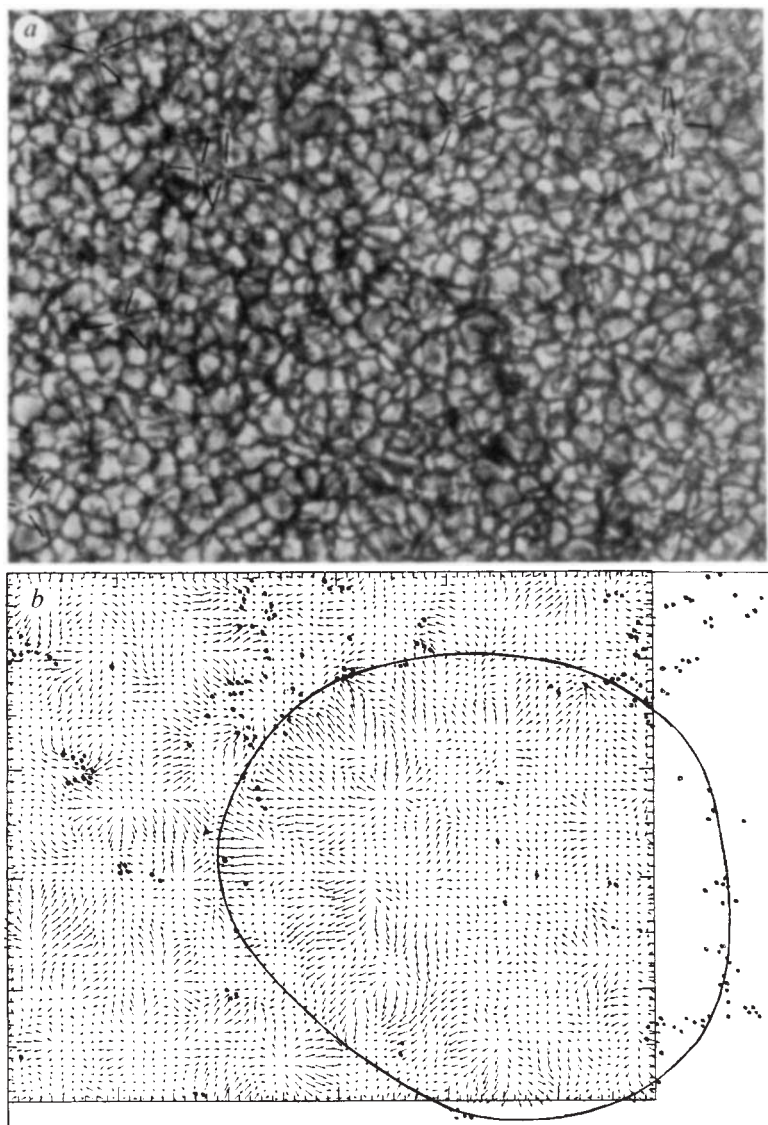


FIG. 1 a, Typical image of solar granulation obtained in the high-quality 3-h data set of 20 September 1988. Area is $70'' \times 50''$. Converging lines point to examples of NBPs (see text). b, Flow vectors in a $58'' \times 48''$ section of the larger area of a (outer rectangle). Line marks supergranule boundary; black dots are loci of NBPs.

correspond to the brightest and darkest features in the grey-scale images. Many of the positive divergence contours (solid lines) represent mesogranules. The dashed-line contours are in regions of converging flows. These form a more connected, linear, pattern than do the positive contours. The supergranule boundary clearly lies largely in such converging regions. The $58'' \times 48''$ area covered by Fig. 2 is exactly cospatial with the vector map of Fig. 1b. Time increases from top to bottom. We have labelled 10 mesogranules in Fig. 2 to describe more easily various aspects of their evolution. Mesogranules 2, 5, 11, 16, 22 and 26 all illustrate our most interesting new result: the advection of mesogranules toward the supergranule boundary. This motion is evident, especially when the mesogranules are located away from the supergranule centre. The translational velocities range from zero to 0.7 km s^{-1} , with a mean velocity of $0.3\text{--}0.4 \text{ km s}^{-1}$.

This is consistent with the horizontal velocity of supergranules obtained from Dopplergrams⁵. Asymmetry in the vector pattern also suggests a mesogranular flow toward the boundary: velocity arrows are longer on one side of the mesogranules, indicating motions towards the supergranular boundary. There are 5–10 well established diverging regions in this portion of the supergranule cell at any time, so that the entire supergranule contains perhaps a dozen such outflows. Visual inspection suggests two types of mesogranular motion: those mesogranules such as 1 that form near the supergranule centre move randomly, whereas those that emerge closer to a supergranule boundary drift towards that boundary; when they reach it, they collapse and disappear. Examples of disappearance are 15, 16 and 22, whereas mesogranules 3, 9 and 11 appear during the run. On average, two of these sources appear or disappear each hour in

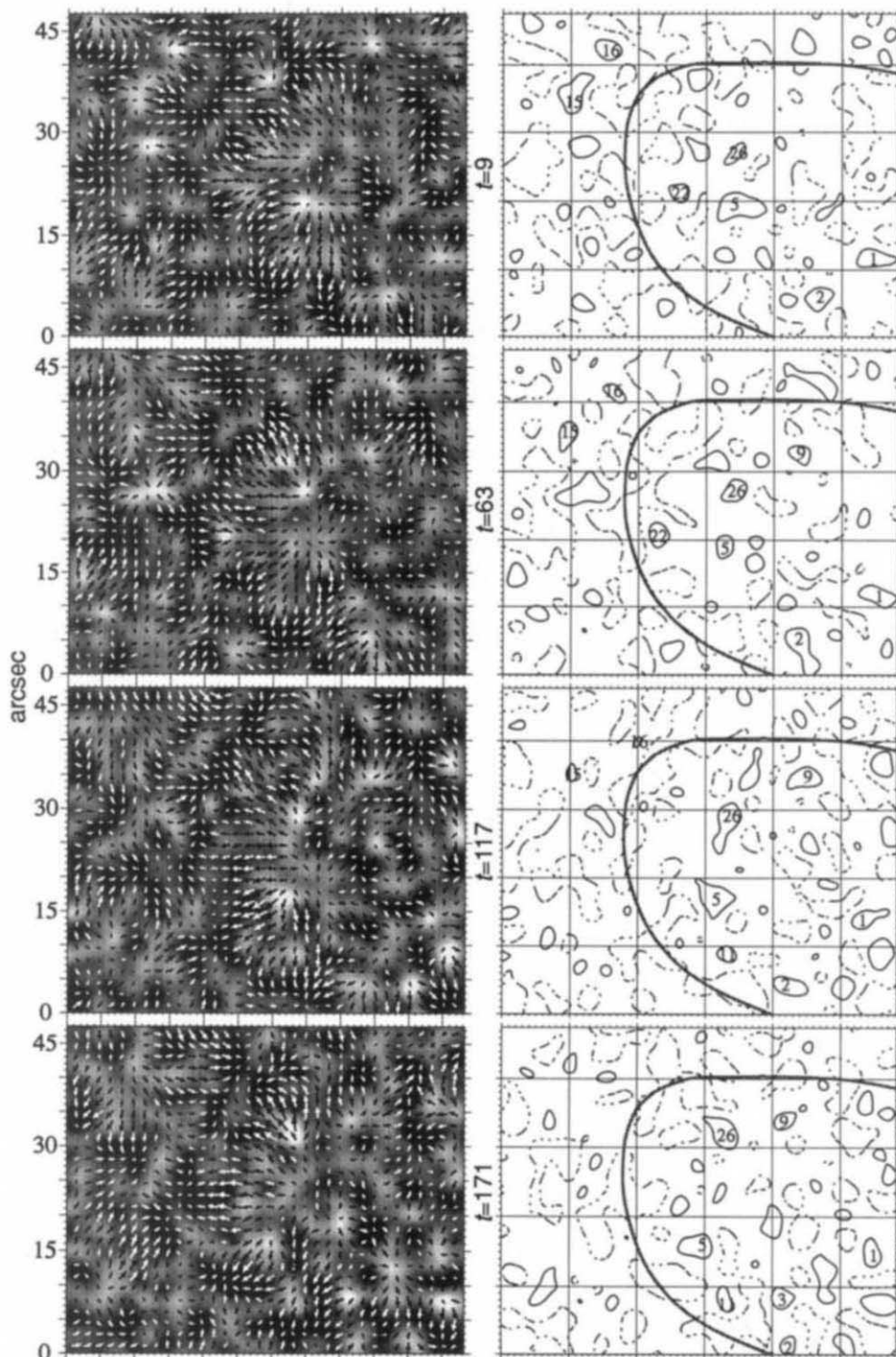


FIG. 2 Left column: velocity vectors superposed on grey-scale images of the flow divergence at 54-min intervals during the sequence. Right column: contours of the divergence include 10 examples of mesogranule evolution, as described in the text. Same $58'' \times 48''$ area as in Fig. 1b; hand-drawn lines again show boundary of supergranule.

the area studied, and four of the labelled mesogranules (1, 2, 5 and 26) are present throughout the run. Results from this small data sample imply a mesogranule lifetime of ~ 3 h. As some mesogranules suffer premature deaths when they collide with the supergranule boundary, their intrinsic lifetimes may be somewhat longer¹⁴. A close look at the four vector images in Fig. 2 shows that the predominant arrow pattern near the supergranule boundary represents a long-lived flow toward that boundary, one that hardly changes in three hours. In fact it is well known⁵ that the supergranule lifetime is about a day or longer. Although we have illustrated our results with only four time steps, the actual analysis involved a detailed study of 78 such images made every two minutes over the 3-h sequence.

The mesogranular pattern is thus more dynamic and irregular than that of supergranules. Part of this difference is due to the shorter lifetimes and smaller sizes of the mesogranules, but most of it probably results from fluctuations of their flow fields due to exploding granules¹⁵ (bright granules that develop a dark centre and then expand radially as a bright annulus until they burst into fragments) within them, horizontal advection by the supergranular flows, and finally their collisions with the supergranule boundary. What, then, is the nature of the mesogranulation? It has been suggested recently¹⁶ that granules and supergranules are the two basic scales in the Sun's convection zone, whereas the mesogranules occur as transient secondary features within the supergranule, similar to convection patterns observed in the laboratory^{17,18} and in theoretical models¹⁷.

Our analysis of the Pic du Midi movie has revealed a hierarchy of advecting convective flow patterns at the surface of the Sun. We have confirmed that granules move in the flow fields of mesogranules and have discovered that the mesogranules, in turn, are advected by the still-larger supergranules. As this pattern of motions evolves with time in a statistically random fashion, it drags along bits of magnetic flux that appear at the Sun's surface. This diffusion process first concentrates the flux into the supergranular network^{5,19}, but as the supergranules themselves evolve over days and weeks, magnetic field will be redistributed over much larger regions of the Sun. In conjunction with the Sun's poleward-moving meridional velocity field, this mechanism may be partially responsible for large-scale magnetic field transport over the Sun during the solar cycle²⁰.

Recent kinematic modelling²¹ of these motions has considerably clarified our understanding of the relative roles of granules, exploding granules, mesogranules and supergranules in the evolution of solar surface flow patterns, but to understand these phenomena in detail will require theoretical modelling of three-dimensional compressible convection, a difficult area in which considerable progress is now being made²²⁻²⁴. The Sun continues to be a unique laboratory for the study of magnetoconvection. □

21. Simon, G. W., Title, A. M. & Weiss, N. O. *Astrophys. J.* **375**, 775-788 (1991).
22. Spruit, H. C., Nordlund, Å. & Title, A. M. *Ann. Rev. Astr. Astrophys.* **8**, 263-301 (1990).
23. Stein, R. F. & Nordlund, Å. *Astrophys. J.* **342**, L95-L98 (1989).
24. Cattaneo, F., Brummell, N. H., Toomre, J., Malagoli, A. & Hurlburt, N. E. *Astrophys. J.* **370**, 282-294 (1991).

ACKNOWLEDGEMENTS. We thank P. Brandt, J. Toomre and N. Weiss for discussion. H.A., R.M., T.R. and J.V. thank the US Air Force Office of Scientific Research and its Window-on-Science Program, the French Ministère des Affaires Étrangères and the Groupe de Recherche du CNRS "Magnétisme dans les Étoiles de Type Solaire" for financial support, and also acknowledge the kind hospitality of the Lockheed Palo Alto Research Laboratory and the National Solar Observatory. This project was partially supported by NASA and the NSF.

The molecular necklace: a rotaxane containing many threaded α -cyclodextrins

Akira Harada, Jun Li & Mikiharu Kamachi

Department of Macromolecular Science, Faculty of Science, Osaka University, Toyonaka, Osaka 560, Japan

THE importance of non-covalent interactions in biological systems motivates much of the current interest in supramolecular assemblies¹. A classic example of a supermolecule is provided by the rotaxanes²⁻⁵, in which a molecular 'rotor' is threaded by a linear 'axle'. Previous examples have included cyclic crown ethers threaded by polymers⁶, paraquat-hydroquinone complexes⁷ and cyclodextrin complexes^{8,9}. We found recently that α -cyclodextrin will form high yields of a crystalline complex with polyethylene glycol (PEG), and suggested that the PEG penetrates the 'beaker-like' tunnel of the cyclodextrin^{10,11}. We report here the preparation of a compound in which several cyclodextrins are threaded on a single PEG chain and are trapped by capping the chain with bulky end groups. This brings a step closer the 'molecular abacus' proposed by Stoddart and coworkers⁷. We call this supramolecular assembly a 'molecular necklace'.

We obtained inclusion complexes of α -cyclodextrin (α -CD) with poly(ethylene glycol) bisamine (PEG-BA) (relative

TABLE 1 NMR spectral data for the product

	Shift (p.p.m.) in DMSO-d ₆ (270 MHz)	Shift (p.p.m.) in D ₂ O + NaOH (270 MHz)
¹ H NMR		
meta H of phenyl	8.87(s, 2H)	8.93(s, 2H)
meta H of phenyl	8.30(d, 2H)	8.17(d, 2H)
ortho H of phenyl	7.26(d, 2H)	7.05(d, 2H)
O(2)H of α -CD	5.63(s, 6H $\times$ 20)	
O(3)H of α -CD	5.48(s, 6H $\times$ 20)	
C(1)H of α -CD	4.80(s, 6H $\times$ 20)	4.74(d, 6H $\times$ 20)
O(6)H of α -CD	4.40(s, 6H $\times$ 20)	
C(3)H, C(6)H,	3.64-3.74	3.59-3.75
C(5)H of α -CD	(m, 24H $\times$ 20)	(m, 24H $\times$ 20)
CH ₂ of PEG	3.51(s, 4H $\times$ 82)	3.50(s, 4H $\times$ 82)
C(2)H and C(4)H of α -CD	3.24-3.29	3.19-3.32
	(m, 12H $\times$ 20)	(m, 12H $\times$ 20)
	Shift (p.p.m.) in DMSO-d ₆ (125.65 MHz)	
¹³ C NMR		
C(1) of α -CD	101.93	
C(4) of α -CD	81.69	
C(3) of α -CD	73.33	
C(2) of α -CD	72.08	
C(5) of α -CD	71.53	
PEG	69.38	
C(6) of α -CD	59.74	

Shifts are given in p.p.m. relative to solvent (DMSO-d₆ is a standard deuterated solvent). Relative intensities are given in parentheses, s, d and m refer to singlet, doublet and multiplet states.

Received 13 September 1991; accepted 14 February 1992.

1. Hale, G. *Astr. Astrophys.* **11**, 407-417 (1892).
2. Deslandres, H. *C. R. Acad. Sci. Paris* **118**, 842-844 (1894).
3. Hart, A. *Mon. Not. R. astr. Soc.* **116**, 38-55 (1956).
4. Leighton, R. B., Noyes, R. W. & Simon, G. W. *Astrophys. J.* **135**, 474-499 (1962).
5. Simon, G. W. & Leighton, R. B. *Astrophys. J.* **140**, 1120-1147 (1964).
6. November, L. J., Toomre, J., Gebbie, K. & Simon, G. W. *Astrophys. J.* **245**, L123-L126 (1981).
7. Title, A. M. *et al. Adv. Space Res.* **6**, 253-262 (1986).
8. Simon, G. W. *et al. Astrophys. J.* **327**, 964-967 (1988).
9. November, L. J. *Appl. Optics* **25**, 392-397 (1986).
10. November, L. J. & Simon, G. W. *Astrophys. J.* **333**, 427-442 (1988).
11. Stenflo, J. O. & Harvey, J. W. *Solar Phys.* **95**, 99-118 (1985).
12. Muller, R. *Solar Phys.* **100**, 237-255 (1985).
13. Title, A. M., Shine, R. A., Tarbell, T. D., Topka, K. P. & Scharmer, G. B. in *Solar Photosphere: Structure, Convection and Magnetic Fields*, IAU Symp. No. 138 (ed. Stenflo, J.) 49-66 (Kluwer, Dordrecht, 1989).
14. November, L. J. *Astrophys. J.* **344**, 494-503 (1989).
15. Rösch, J. & Hugon, M. *C. R. Acad. Sci. Paris* **249**, 625-627 (1959).
16. Simon, G. W. & Weiss, N. O. *Mon. Not. R. astr. Soc.* **252**, 1p-5p (1991).
17. Busse, F. H. in *Hydrodynamic Instabilities and the Transition to Turbulence* (ed. Swinney, H. L. & Gollub, J. P.) 97-137 (Springer, Berlin) 1981.
18. Zocchi, G., Moses, E. & Libchaber, A. *Physica A* **166**, 387-407 (1990).
19. Title, A. M. *et al. Astrophys. J.* **336**, 475-494 (1989).
20. Wang, Y.-M., Nash, A. G. & Sheeley, N. R. Jr *Science* **245**, 712-718 (1989).

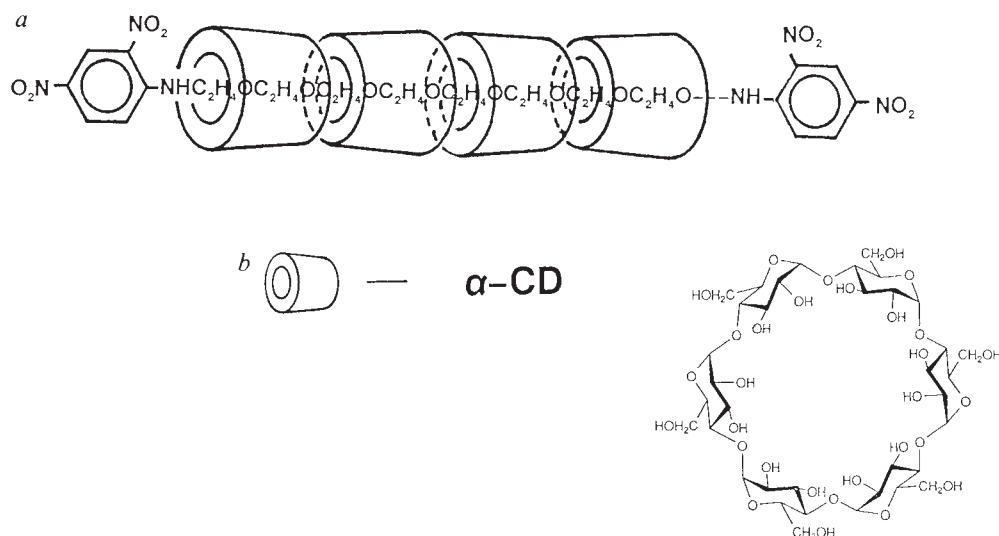


FIG. 1 a, Molecular necklace prepared from complexes of poly(ethylene glycol) bisamine with α -cyclodextrins by capping the complexes with 2,4-dinitrofluorobenzene. b, α -Cyclodextrin.

molecular mass, $M_r = 3,350$) by adding an aqueous solution of PEG-BA to a saturated aqueous solution of α -CD at room temperature, using a method similar to that used to prepare complexes of α -CD and PEG¹⁰. The resulting complex was collected and dried. We added to the complex an excess (46 equivalents) of 2,4-dinitrofluorobenzene, which is bulky enough to prevent dethreading¹², together with dimethylformamide, and stirred the mixture at room temperature overnight (Fig. 1). The reaction was stopped in excess ether to precipitate the product. The precipitate was washed with ether (three times) to remove unreacted 2,4-dinitrofluorobenzene and with dimethylformamide to remove free α -CD, PEG-BA, and dinitrophenyl derivatives. The residue was dissolved in dimethylsulphoxide (DMSO) and washed with water to remove unreacted α -CD, PEG-BA and water-soluble dinitrophenyl derivatives. The product was collected, washed with ether and dried (yield, 60%). Finally, the product was purified by column chromatography on Sephadex G-50 by using DMSO as solvent. The product was found to be pure and contained no free α -CD, PEG-BA or dinitrophenyl derivatives.

The product is insoluble in water and dimethylformamide although each component, α -CD, PEG-BA, bis(2,4-dinitrophenyl)-PEG (hereafter called PEG-DNB₂), and even α -CD-PEG-BA complexes are soluble in water. The product is, however, soluble in DMSO and 0.1M NaOH. The hydroxyl groups of α -CD ($pK_a = 12$) may be ionized at this pH, so that α -CD becomes soluble in an aqueous medium. Neutralizing this solution by adding 0.1M HCl instantly produced a precipitate. This reaction is reversible. The structure of the product was confirmed by elemental analysis, infrared and ultraviolet spectroscopy, X-ray diffraction, and proton and ¹³C nuclear magnetic resonance (NMR). Combustion analysis gave the following results: found (calculated) for C₈₈₈H_{1,520}N₆O₆₈₅·(H₂O)₂₀: C, 44.86(45.18); H, 6.69(6.66); N, 0.38(0.36). Proton and ¹³C NMR spectra of the product are consistent with a compound that contains both components α -CD and PEG-DNB₂, although some broadening was observed. Table 1 shows NMR data for the product. Gel chromatography on Sephadex G-50 (exclusion limit 30,000) using DMSO as eluent shows a single peak close to the void volume. Figure 2 is a chromatogram of a mixture of product (3 mg), α -CD (30 mg) and PEG-DNB₂. The chromatogram shows three peaks. The first peak, which could be detected by both ultraviolet spectroscopy (316.5 nm) and optical rotation, was identified as the product. The second, which could be detected only by ultraviolet, was identified as PEG-DNB₂. The third, which could be detected only by optical rotation, was identified as α -CD. The average molecular mass

was determined by ¹H NMR to be 23,200, which indicates that ~20 α -CD units are captured in a polymer chain, and by ultraviolet spectroscopy to be 26,400, indicating 23 α -CD units in a single molecule. We consider these to be the same within experimental error. X-ray powder diffraction shows that the complexes are crystalline, and the patterns are similar to those of complexes of α -CD with valeric acid or octanol, which differ from those of α -CD complexes with small molecules, such as propionic acid or propanol. These results indicate that the product is isomorphous with those of channel-type structure.

At present we have not determined how adjacent cyclodextrins are oriented. But the fact that the product is insoluble in water but soluble in 0.1M NaOH shows that strong hydrogen bonds exist between α -CDs, suggesting that they alternate as shown in Fig. 1. Preliminary results on thermodynamics (microcalorimetry) for complex formation show that the complexation process is exothermic, indicating hydrogen-bond formation between α -CDs (H. Suga *et al.*, manuscript in preparation).

PEG cannot form complexes with α -CD if each end carries a bulky substituent, such as the 3,5-dinitrobenzoyl group or 2,4-dinitrophenyl group, which does not pass through the α -CD cavity. Therefore we conclude that, on average, 20–23 α -CDs are entrapped on a single molecule, like beads on an abacus. This is an example of a rotaxane which includes many 'rotors' in a single molecule. Such molecular organization is important

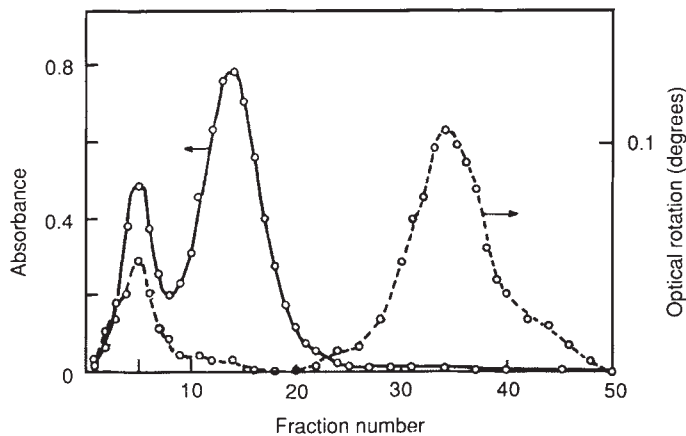


FIG. 2 Chromatography of the mixture of the product (3 mg), α -CD (30 mg) and PEG-DNB₂ on a Sephadex G-50 column in DMSO (1.7 × 70 cm). Absorbance was measured at 316.5 nm, 1.5 ml per fraction. Optical rotation was measured at 589 nm with cell length 10 cm.

not only in biological systems and chemical processes, but also in the creation of molecular devices.

After this manuscript had been submitted, Isnin and Kaifer¹³ reported the synthesis of a zwitterionic rotaxane containing a single cyclodextrin molecule. □

Received 9 September 1991; accepted 12 February 1992.

1. Lehn, J.-M. *Angew. Chem. Int. Ed. Engl.* **29**, 1304–1319 (1990).
2. Agam, G., Graiver, D. & Zilkha, A. *J. Am. chem. Soc.* **98**, 5206–5216 (1975).
3. Agam, G. & Zilkha, A. *J. Am. chem. Soc.* **98**, 5214–5216 (1976).
4. Ogino, H. *J. Am. chem. Soc.* **103**, 1303–1304 (1981).
5. Ogino, H. & Ohata, K. *Inorg. Chem.* **23**, 3312–3316 (1984).
6. Gibson, H. W. et al. *Makromolec. Chem., Makromolec. Symp.* **42/43**, 395 (Hüthig & Werf, Basel, 1991).
7. Ashton, P. R. et al. *Angew. Chem. Int. Ed. Engl.* **30**, 1042–1045 (1991).
8. Manka, J. S. & Lawrence, D. S. *J. Am. chem. Soc.* **112**, 2440–2442 (1990).
9. Rao, T. V. S. & Lawrence, D. S. *J. Am. chem. Soc.* **112**, 3614–3615 (1990).
10. Harada, A. & Kamachi, M. *Macromolecules* **23**, 2821–2823 (1990).
11. Harada, A. & Kamachi, M. *J. chem. Soc., Chem. Commun.* 1322–1323 (1990).
12. Bergeron, R. J., Channing, M. A., Gibelly, G. J. & Pillor, D. M. *J. Am. chem. Soc.* **99**, 5146–5151 (1977).
13. Isnin, R., Kaifer, A. E. *J. Am. chem. Soc.* **113**, 8188–8190 (1991).

Imaging the ocean with ambient noise

Michael J. Buckingham^{*†}, Broderick V. Berkhout^{*}
& Stewart A. L. Glegg[‡]

^{*} Marine Physical Laboratory, Scripps Institution of Oceanography, La Jolla, California 92093-0213, USA

[‡] Department of Ocean Engineering, Florida Atlantic University, Boca Raton, Florida 33431, USA

FOR many years, the principal means of probing the ocean using sound has been through the use of 'active' or 'passive' techniques. With an active system, an object is illuminated by a pulse of sound and its presence inferred from the echo it produces, whereas the passive approach involves simply listening for the sound that the object itself emits. Here we report a new method of using sound in the ocean, which is neither passive nor active. It relies on the naturally occurring, incoherent ambient noise field in the ocean—which can be thought of as 'acoustic daylight'—as the sole source of acoustic illumination. By focusing the sound scattered by an object immersed in the noise field, it should be possible to produce a visual image of the object on a television monitor. We have tested this concept by conducting a simple experiment in the ocean, with a parabolic reflector acting as an acoustic lens, and our results confirm that objects illuminated only by ambient noise can indeed be 'seen' at frequencies between 5 and 50 kHz.

The ocean is an acoustically noisy environment¹. Sources of the ambient noise field include a variety of natural mechanisms², many of which are associated with breaking surface waves³ and other surface phenomena. Over the frequency range between 1 and 50 kHz, the power spectral density of sea-surface noise shows a negative gradient, varying with frequency f as f^{-n} , where $n \approx 2$. Surface-generated noise takes the form of an incoherent, propagating radiation field, with energy travelling in all directions, and in this sense is analogous to daylight in the atmosphere. Above 50 kHz or so, depending on sea state, thermal noise arising from brownian motion of the water molecules dominates the ambient noise spectrum⁴. Thermal noise is a localized, non-radiating, microscopic phenomenon, which shows a spectrum with positive gradient, varying with frequency as f^2 . Figure 1 illustrates both the radiating, surface-generated noise and the thermal noise components of the ambient noise spectrum.

Recent experiments⁵ have shown that geo-acoustic information about the ocean environment can be extracted from the

ambient noise. Here we report the use of ambient noise to form pictorial images of objects in the ocean. We can draw an analogy between the natural optical (daylight) field in the atmosphere and the radiating ambient noise in the ocean: both fields consist of random, incoherent radiation propagating in all directions. For this reason the concept of imaging with ambient noise, which was proposed by one of us (M.J.B.) several years ago, has been designated 'acoustic daylight'.

An object in the ambient noise field modifies the field by scattering acoustic energy in all directions. This scattered radiation can be focused onto an image plane by an acoustic lens (which could be an acoustic refractor, reflector or phased array), allowing, after appropriate signal processing, a pictorial image of the object to be displayed on a television monitor. In principle, the screen image could move and could have computer-constructed colour, corresponding to the 'acoustic colour' (the frequency-dependent acoustic albedo) of the object space. In effect, the acoustic daylight scheme consists of an acoustic 'video camera', which relies on illumination from the natural sound field in the ocean to produce the final picture.

Although the analogy between daylight and ambient noise provides a useful introduction to the concept of acoustic daylight, it obviously cannot be pursued too far. For instance, the acoustic wavelength in the ocean at a frequency of 50 kHz is 3 cm, compared with 5×10^{-5} cm for the wavelength of green light in the atmosphere. The pupil of the human eye is $\sim 20,000$ optical wavelengths in diameter, which accounts for the excellent angular resolution of human vision. To achieve similar resolution with an acoustic lens working at 50 kHz would require an aperture of 600 m. Practical considerations dictate that a much smaller aperture be used in practice, yielding a corresponding reduction in angular resolution. Nevertheless, satisfactory images could still be obtained, especially if image enhancement techniques similar to those developed for synthetic-aperture radar and infrared remote sensing imagery were applied.

Another factor that will limit the performance of an acoustic daylight imaging system is sound absorption in the ocean. At 20 kHz the volume attenuation is ~ 3 dB km⁻¹ (ref. 1), suggesting that the range achievable with the system is likely to be less than 1 km, although an improvement could be obtained by working at lower frequencies provided the reduction in angular resolution could be tolerated. For the moment, a reasonable operating frequency range for the acoustic daylight system is taken to be 5–50 kHz, with the lower limit set (rather arbitrarily) by angular resolution considerations and the upper limit determined by the onset of thermal noise (Fig. 1). As thermal noise is molecular in origin, it contains no information about the object and hence cannot be used for imaging.

Imaging with incoherent, natural sound is distinct from both

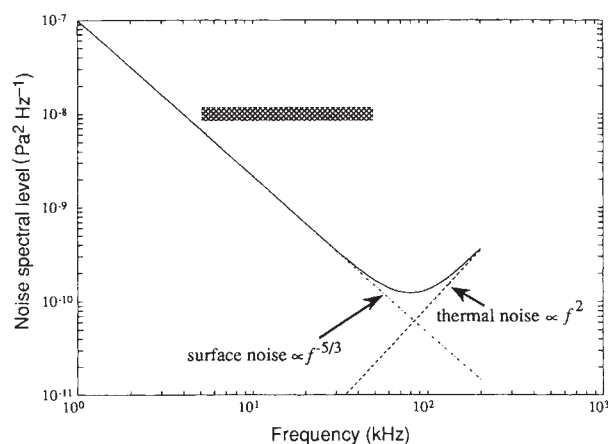


FIG. 1 Nominal ambient noise spectrum (solid line) in the ocean. The cross-hatched bar depicts the frequency range of interest in the acoustic daylight imaging experiment.

[†] Also at the Institute of Sound and Vibration Research, The University, Southampton SO9 5NH, UK.

passive and active detection, which are the conventional acoustic methods used to probe the ocean. Potentially, the new technique offers several advantages. As it does not require the transmission of acoustic energy (unlike active systems), it is attractive in connection with submarine operations. For example, mounted on the bow of a submarine, an acoustic daylight 'camera' could act as an aid to navigation by providing forward vision, a facility which would be particularly useful under the Arctic ice cap, where ice keels may protrude 40 m or more below the surface. Acoustic daylight might also be useful for undersea detection, as quiet or silent objects, which are undetectable by passive systems, are visible with the incoherent imaging technique.

Other possible uses include subsurface monitoring of oil rigs, surveillance of harbour entrances and bottom surveying, all of which would benefit from the fact that no sound transmission is required. A further desirable feature of the technique is that 'unwanted' noise sources, including reverberation and multipath, which degrade active and passive system performance, actually enhance the acoustic daylight image by providing better acoustic illumination. And perhaps it is worth emphasizing that the final acoustic-daylight display would be a pictorial image on a screen, unlike the outputs of active and passive systems, which can be interpreted only by trained operators.

To test the acoustic daylight concept, we recently did a simple experiment off the end of Scripps pier in which a parabolic reflector of diameter 1.22 m, with a neoprene rubber, pressure-release surface, was used as the acoustic lens. A low-noise hydrophone, shielded at the rear to block direct radiation, was mounted at the focus of the parabolic dish. The water depth at the end of the pier is ~ 7 m, and the weather conditions during the 3-day experiment were calm (sea-state 1 or less).

Three rectangular targets, consisting of thick plywood board with neoprene rubber facing, were the objects to be 'imaged'. The targets were placed in the main lobe of the reflector, either broadside-on (the 'on' position) or edge-on (the 'off' position) to the parabolic dish. Two orientations of the beam were tried:

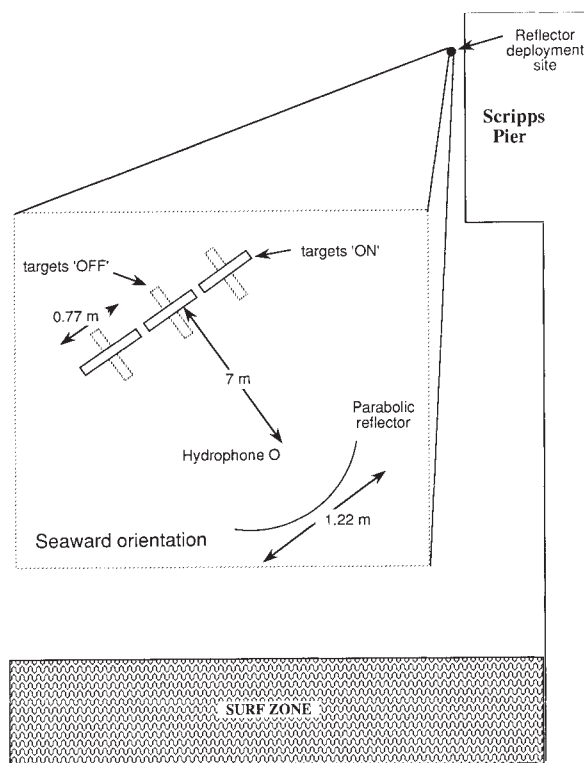


FIG. 2 Plan view of the seaward configuration of the parabolic reflector and targets off Scripps pier. The deployment site is ~ 200 m out from the shore. (Drawing not to scale).

obliquely seaward, as illustrated in Fig. 2, and obliquely shoreward. As the results obtained with the seaward and shoreward configurations are qualitatively similar, only the former are reported here.

The three targets, each of which is 0.9 m high and 0.77 m wide, were placed ~ 7 m from the reflector. With this arrangement, the total target width was the same as the width of the main lobe of the reflector, as measured at the -6 dB points, at a frequency of 9 kHz. Throughout most of the 5–50 kHz frequency band of interest, therefore, the targets filled the beam. Noise spectra taken over the frequency range 5–50 kHz for the seaward orientation of the beam are shown in Fig. 3a. Both spectra are averages of several hundred spectra obtained from time-series data falling within a window of 2 s duration. The spectral resolution is 100 Hz.

The solid line in Fig. 3a represents the noise level with the targets 'on' and the broken line with the targets 'off'. The noise level across the band is ~ 4 dB higher when the targets are 'on' than when they are 'off'. Such behaviour was observed consistently throughout the duration of the experiment, and it is also consistent with the results of analogous experiments⁶ conducted in air, in a hard-walled laboratory, with sound fields generated by loudspeakers. Figure 3b, showing the difference between the 'on' and 'off' spectra in Fig. 3a, indicates clearly that we can 'see' the targets when they are illuminated solely by the natural ambient noise field in the ocean.

The observed difference signal of nominally 4 dB is believed to occur because the noise around Scripps pier shows some horizontal directionality, being generated largely by the pier itself. (The pilings of the pier are encrusted with molluscs, and passing waves create an audible signal in the atmosphere.) Our experimental arrangement was evidently somewhat akin to taking a photograph with the Sun behind the camera. A second factor influencing the level of the difference signal is the reflectivity of the targets, which, in our case, were perfect acoustic

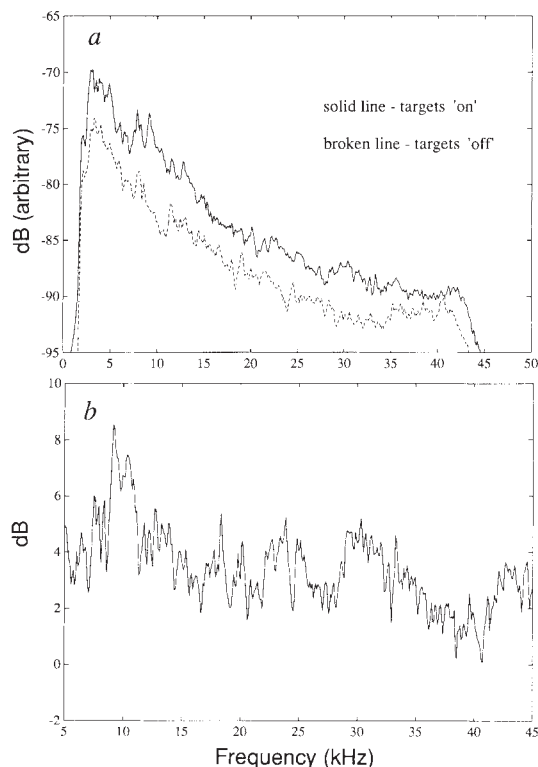


FIG. 3 a, Noise spectra obtained with the beam of the parabolic reflector orientated seaward and the targets at a range of 7 m. (The rapid roll-off above 43 kHz is an effect of anti-alias filtering.) b, Difference spectrum, illustrating the nominally uniform separation between the 'on' and 'off' spectra over most of the frequency band of interest.

reflectors; had they been absorbers of sound, they would have left a 'hole' in the noise field and the difference signal would have been zero or negative. A negative signal can form the basis of an image, a 'silhouette', the important criterion being that there should be an acoustic contrast between the object and its background.

Although this first acoustic daylight experiment in the ocean was successful, the observed spectrum is a far cry from a pictorial image on a television monitor. In effect, we have generated just one pixel of an image, corresponding to the single beam of the parabolic reflector. More pixels require more beams. These could be produced with a volumetric phased array, although other possibilities involving reflectors, either parabolic or spherical, and refractive acoustic lenses are also being considered. □

Received 6 November 1991; accepted 11 February 1992.

1. Urlick, R. J. *Principles of Underwater Sound* (McGraw-Hill, New York, 1975).
2. Kerman, B. *Sea Surface Sound* (Kluwer, Dordrecht, 1988).
3. Melville, W. K., Loewen, M. R., Felizardo, F. C., Jessup, A. T. & Buckingham, M. J. *Nature* **336**, 54–56 (1988).
4. Mellen, R. H. *J. acoust. Soc. Am.* **24**, 478–480 (1952).
5. Buckingham, M. J. & Jones, S. A. *J. acoust. Soc. Am.* **81**, 938–946 (1987).
6. Glegg, S. A. L. & Buckingham, M. J. *Florida Atlantic Univ. Rep.* (July 1990).

ACKNOWLEDGEMENTS. We thank E. Slater and L. Green for the loan of the parabolic reflector and G. Deane for guidance on design of circuit elements. This work was supported by the Office of Naval Research.

Analytical model for solidification of the Earth's core

Bruce A. Buffett, Herbert E. Huppert, John R. Lister & Andrew W. Woods

Institute of Theoretical Geophysics, Department of Applied Mathematics and Theoretical Physics and Department of Earth Sciences, University of Cambridge, Cambridge CB3 9EW, UK

THE Earth's solid inner core is generally thought to have formed by gradual solidification of the liquid core as the Earth cooled^{1–3}. To elucidate the relative importance of the various physical effects on the thermal evolution of the core, we have developed an analytical model based on global heat conservation, which describes the cooling of the vigorously convecting, fluid outer core and the concomitant growth of the inner core. We obtain a simple form for the evolution of the inner-core radius which allows the consequences of changes to the model's input parameters to be readily assessed. For most of this evolution, inner-core growth is controlled primarily by the heat capacity of the outer core and the history of the heat flux into the base of the mantle. Heat sources associated with solidification of the inner core, including the release of latent heat and gravitational energy, have a secondary role but become more important towards the end of solidification. Using current seismic estimates of compositional changes at the surface of the inner core, we conclude that the compositional and thermal buoyancy fluxes in the outer core are comparable.

The energetics of the cooling core have been studied previously^{4–7} to estimate the energy available to power the geodynamo through the release of latent heat on solidification⁸ and through the associated generation of compositional buoyancy⁹ by the exclusion of light elements from the iron-rich inner core. Other studies^{10–12} have attempted to construct thermal histories of the Earth's core using parameterized models of convection. The latter have all involved extensive numerical calculations, in contrast to the analytical model developed here.

In our model we assume that compositional and thermal buoyancy forces drive vigorous convection in the liquid outer core, and that the liquid core is well mixed with an adiabatic temperature profile. The potential temperature in the liquid core,

which is defined as the temperature after subtracting the adiabatic variation, is therefore spatially uniform and slowly decreases with time (Fig. 1). We also assume that the surface of the inner core is in thermodynamic equilibrium with the surrounding liquid. With these two assumptions, the temperature through the outer core is uniquely determined by the solidification temperature, $T_s(p)$, which we treat as a function of pressure p only and hence implicitly of depth. The solidification temperature may also be a function of the slowly varying composition. At present, the details of the phase diagram for the iron alloy constituting the core are poorly constrained, but simple theoretical models suggest that the effect of compositional variations on T_s is small⁷.

The heat extracted from the cooling outer core is transported through the overlying mantle to the surface of the Earth by thermal convection. Convective motions in the mantle¹³ are characterized by effective viscosities of 10^{21} – 10^{23} Pa s (refs 14–16). The cooling timescale associated with mantle convection is therefore much greater than that characteristic of convection in the outer core, where the viscosity is less than 5×10^3 Pa s (ref. 17). Thus the relatively sluggish and more massive mantle controls the cooling of the core by regulating the heat flux $f_m(t)$ across the core–mantle boundary. The magnitude and time dependence of $f_m(t)$ will depend on the details of mantle convection, which include the distribution of radioactive isotopes and any compositional or rheological layering in the mantle. As many such details are unknown, we treat $f_m(t)$ as a prescribed parameter and focus on the evolution of the core. Different models of mantle convection will give rise to different values of $f_m(t)$, and these may be readily incorporated into our model.

The model is based on global heat conservation, which equates the heat lost from the core liquid plus the heat produced by growth of the inner core to the net heat flux from the outer core. The main sources of heat associated with the growth of the inner core are the latent heat released by solidification and the loss of gravitational energy due to the solidification of an iron-rich inner core. This gravitational energy is converted to heat by buoyancy-driven motions in the core. Further gravitational energy is released by adiabatic compression and thermal contraction of the Earth, but detailed numerical calculations (B.A.B., H.E.H., J.R.L. and A.W.W., manuscript in preparation) show that these additional effects are less important, and they are not considered further here. Radioactive heat sources in the core¹⁸ could be explicitly included in the model, but such heat

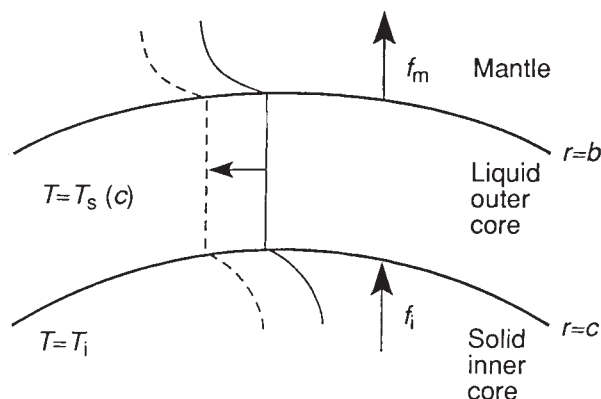


FIG. 1 Schematic profile of the potential temperature (the temperature after subtracting the adiabatic variation) in the slowly cooling core of the Earth. The solid line represents the temperature in the core and lower mantle at some instant. The dashed line represents the subsequent evolution of temperature as heat is continuously extracted from the core. Heat fluxes f_m and f_i pertain to the core–mantle and inner-core boundaries respectively. The solidification temperature $T_s(c)$ is defined at the inner core boundary, and $T_i(r, t)$ is the temperature within the inner core.

sources can be absorbed into f_m . Thus, in spherical geometry, we express the global heat balance as

$$\frac{4\pi}{3}(b^3 - c^3)C_p \frac{dT_s}{dt} - 4\pi c^2(L + B) \frac{dc}{dt} = 4\pi c^2 f_i(t) - 4\pi b^2 f_m(t) \quad (1)$$

where C_p and L are the specific heat and latent heat per unit volume, and c and b are the inner and outer radii of the liquid core. The release of gravitational energy per unit volume B is due to the compositional buoyancy flux at the inner-core boundary. The change in the gravitational energy that results from the rearrangement of mass within the core can be approximated by²

$$B = \Delta\rho g(b) b \left[\frac{3}{10} - \frac{1}{2}(c/b)^2 \right] \quad (2)$$

where $\Delta\rho$ is the compositional density jump across the inner-core boundary and g is the local acceleration due to gravity. The conductive heat flux f_i from the inner core is obtained from

$$f_i(t) = -k \nabla T_i(c, t) \quad (3)$$

where $T_i(r, t)$ is the temperature within the solid inner core and k is the thermal conductivity. Time t is measured from the instant at which the temperature of the core first falls below the solidification temperature at the centre of the Earth and the inner core begins to grow.

We circumvent the need for an explicit solution for the temperature in the interior of the inner core by considering two bounding cases for f_i , and hence for T_i . These two endmembers correspond physically to the limits of: (1) a perfectly insulating inner core, $k \rightarrow 0$; and (2) a perfectly conducting inner core, $k \rightarrow \infty$. In the former case, no heat is conducted from the inner core and $f_i = 0$. In the latter case, heat is readily conducted from the inner core, thus maintaining an isothermal temperature field within the solid given by the solidification temperature T_s at the interface. Consideration of the heat content of the inner core then shows that $f_i \rightarrow -\frac{1}{3}cC_p(dT_s/dt)$ as $k \rightarrow \infty$.

The equation governing the radius $c(t)$ of the inner core is obtained by making the radial dependence of $T_s(p)$ explicit in equation (1). The relationship between p and r is determined by the hydrostatic equation $\partial p / \partial r = -\rho g$ which dominates the radial momentum balance of the Earth, where ρ is the mass density and g can be obtained from Newton's law of gravitation. In our calculations we neglect the known radial and temporal

TABLE 1 Physical properties of the Earth's core

Parameter		
Thermal expansion coefficient ²³	α	$7 \times 10^{-6} \text{ } ^\circ\text{C}^{-1}$
Latent heat ²⁴	L	$6.0 \times 10^9 \text{ J m}^{-3}$
Specific heat capacity ²³	C_p	$6.7 \times 10^6 \text{ J } ^\circ\text{C}^{-1} \text{ m}^{-3}$
Solidification profile ²⁴	$\partial T_s / \partial p$	$7 \times 10^{-9} \text{ } ^\circ\text{C Pa}^{-1}$
Mean density ²²	ρ	$1.2 \times 10^4 \text{ kg m}^{-3}$
Density jump at inner-core boundary ²²	$\Delta\rho$	600 kg m^{-3}
Outer-core radius ²²	b	3,480 km
Current inner-core radius ²²	c	1,221 km
Gravitational constant	G	$6.67 \times 10^{-11} \text{ m}^3 \text{ kg}^{-1} \text{ s}^{-2}$
Current radius ratio	c/b	0.35
Model constant	$\mathcal{M}$	$1.3 \times 10^{16} \text{ J m}^{-2}$
Latent heat parameter	$\mathcal{L}$	0.5
Gravitational energy parameter	$\mathcal{B}$	0.6

variations in density because their effects are small compared with both the leading order radial variation and the uncertainties in $T_s(p)$. Over a limited range of pressure, $T_s(p(r))$ can be approximated using a Taylor series expansion about $p(0)$. Hence to first order, the radial dependence of T_s is

$$T_s(r) = T_s(0) - \frac{2\pi}{3} G \rho^2 r^2 \frac{\partial T_s}{\partial p} \quad (4)$$

where G is the gravitational constant.

Equation (4) relates the temperature of the outer core to the radius of the inner core. Substituting (4) into (1), we obtain equations for the growth of the inner core which can be integrated directly in the two limiting cases, $k \rightarrow 0$ and $k \rightarrow \infty$. For these two cases, $c(t)$ is given by

$$\left(\frac{c}{b}\right)^2 \left[1 - \frac{2}{5} \left(\frac{c}{b}\right)^3 \right] + \mathcal{L} \left(\frac{c}{b}\right)^3 + \mathcal{B} \left(\frac{c}{b}\right)^3 \left[1 - \left(\frac{c}{b}\right)^2 \right] = \frac{1}{\mathcal{M}} \int_0^t f_m(\tau) d\tau \quad (5)$$

and

$$\left(\frac{c}{b}\right)^2 + \mathcal{L} \left(\frac{c}{b}\right)^3 + \mathcal{B} \left(\frac{c}{b}\right)^3 \left[1 - \left(\frac{c}{b}\right)^2 \right] = \frac{1}{\mathcal{M}} \int_0^t f_m(\tau) d\tau \quad (6)$$

respectively, where

$$\mathcal{M} = \frac{2\pi}{9} b^3 C_p G \rho^2 \frac{\partial T_s}{\partial p}, \quad \mathcal{L} = \frac{1}{3} \frac{Lb}{\mathcal{M}}, \quad \mathcal{B} = \frac{1}{10} \frac{\Delta\rho g(b)b^2}{\mathcal{M}} \quad (7)$$

The parameter $\mathcal{M}$ is a measure of the heat that must be extracted to cool the entire liquid core to its solidification temperature. $\mathcal{L}$ and $\mathcal{B}$ represent the importance of the latent heat and the gravitational energy loss, respectively, relative to the heat due to cooling.

One of the most important features of our analytical model is the insight it offers into the processes that dominate the growth of the solid inner core. We can deduce immediately that differences between the two extreme models of a perfectly insulating and perfectly conducting inner core are negligible when $(c/b)^3$ is small. This condition is satisfied over most of the evolution of the core, and so the rate at which heat is lost from the inner core has an insignificant effect on the cooling of the core as a whole. Our analysis also indicates that the effects of latent heat release and loss of gravitational energy are a factor c/b smaller than the heat extracted by cooling the core. For the parameter values given in Table 1, we find that $c/b = 0.35$, $\mathcal{L} = 0.5$ and $\mathcal{B} = 0.6$. This shows that the growth of the inner core is currently controlled by a balance between the heat flux into the base of the mantle and the decrease in the heat content of the cooling outer core. With this approximation, we obtain the following

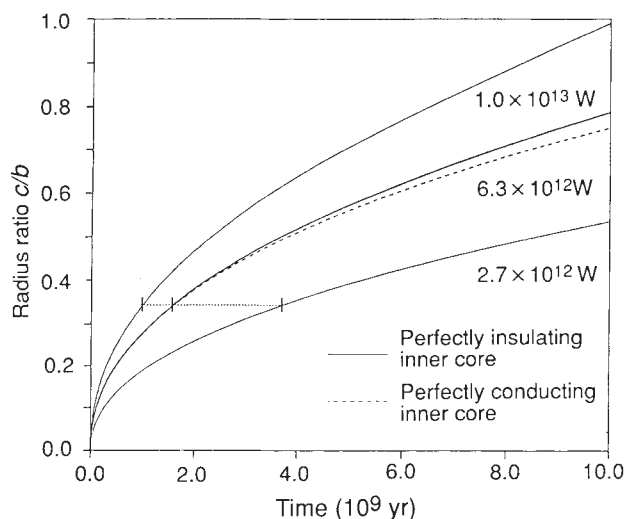


FIG. 2 Growth of the inner core measured from the time at which solidification begins. Three possible solutions are presented using a range of estimates for the net heat flux $4\pi b^2 f_m(t)$ across the core-mantle boundary (see text). The horizontal dotted line indicates the current radius of the inner core. Before $t=0$, the core of the Earth was entirely liquid.

simple expression for the radius of the inner core

$$c(t) = b \left[\int_0^t f_m(\tau) d\tau / M \right]^{1/2} \quad (8)$$

Specific predictions of $c(t)$ from our model require an estimate of $f_m(t)$. A plausible upper bound on f_m can be inferred from the heat flux measured at the Earth's surface. Given chondritic abundances of radioactive isotopes¹⁹ in the upper mantle, it is unlikely that the total heat flux across the core-mantle boundary exceeds 10^{13} W. It is also unlikely, although not impossible, that the total heat flux is less than that conducted up the adiabat, which is estimated¹² to be 2.7×10^{12} W. As an illustrative example, we consider these two bounds on f_m , along with the midpoint value 6.3×10^{12} W as a typical estimate. Using the parameter values in Table 1, we obtain the growth histories shown in Fig. 2. For the typical value of f_m , we see that the inner core grows to its present radius in 1.6×10^9 years, a value similar to those obtained previously by Stevenson *et al.*¹⁰. The values calculated from the bounds on f_m are 1.0×10^9 and 3.6×10^9 years. These ages span a wide interval, but are roughly consistent with palaeomagnetic data²⁰, which are sometimes used to constrain the age of the inner core under the assumption that the onset of solidification is associated with a rapid increase in the strength of the magnetic field.

The thermal buoyancy flux associated with the cooling of the liquid core depends on the amount by which f_m exceeds the heat flux f_a conducted down the adiabatic gradient. For our typical value of f_m , the thermal buoyancy flux caused by the downwelling of cold dense fluid at the core-mantle boundary dominates that associated with latent heat release at the inner-core boundary. But an additional buoyancy flux of compositional origin is produced at the inner-core boundary by the release of light, residual fluid on solidification. The relative importance of the thermal and compositional contributions can be assessed using our model once $\Delta\rho$ and f_m are known. Seismic estimates of the relative density jump $\Delta\rho/\rho$ vary from 2 to 12% (ref. 21), with typical values of 5% (ref. 22). The actual compositional density jump, however, may be less than these seismic estimates if there is any volume change on solidification. In terms of $\Delta\rho$, the compositional buoyancy flux is $4\pi\Delta\rho g(c)c^2(dc/dt)$. For comparison, the thermal buoyancy flux is given by $4\pi\alpha\rho g(b)b^2(f_m - f_a)$, where α is the coefficient of thermal expansion. From the range of possible values of f_m , we estimate that these two fluxes are roughly comparable if the seismic estimates of the density jump across the inner core boundary represent a compositional change. □

Parent-offspring conflict and the recruitment of helpers among bee-eaters

Stephen T. Emlen & Peter H. Wrege

Section of Neurobiology and Behavior, Cornell University, Ithaca, New York 14853, USA

GENETIC conflicts of interest are to be expected between individuals in any non-clonal society¹⁻⁵. One well studied form of conflict is that between parents and their offspring over the amount of parental care provided to the offspring³. A very different manifestation of parent-offspring conflict may occur in certain cooperatively breeding species in which parents (breeders) are assisted in the rearing of young by their grown offspring (helpers)⁶⁻⁸. If helpers have a sufficiently large effect on reproductive success, breeders will enhance their own inclusive fitness more by retaining their offspring as helpers than by allowing them to reproduce on their own^{4,5,9}. We report here that older male white-fronted bee-eaters (typically fathers) actively disrupt the breeding attempts of their sons, and that such harassment frequently leads to the sons joining as helpers at the nest of the harassing father. Calculation of fitness costs and benefits to the various participants helps to clarify both why parents engage in such 'recruitment' behaviour and why sons frequently do not resist.

White-fronted bee-eaters (*Merops bullockoides*) are common birds of the savannas of East and Central Africa. They breed both gregariously (in colonies averaging 200 individuals) and cooperatively (with half of all nesting attempts including non-breeding adult helpers)¹⁰. Young remain with their parental group until they pair at one to two years of age, when females (but not males) disperse to join the group of their mate¹¹. The resulting social organization is one of patrilocal extended family groups (clans). Each clan occupies a stable feeding territory, but commutes to roost and nest at the colony site^{11,12}.

During the breeding season, several pairs in each extended family may initiate breeding. Some individuals suffer harassment from other members of their clan and many initial reproductive attempts are abandoned. We examined the frequency, pattern and consequences of these harassments to test the hypothesis that such interference constitutes a form of parental manipulation to recruit grown offspring into becoming helpers.

We collected observations ad lib during 5 and 1 breeding seasons, respectively, from two individually marked and genealogically known populations of white-fronted bee-eaters at Lake Nakuru National Park, Kenya. We defined harassment as involving one or more of the following. (1) Prolonged aggressive chasing of a resident bird on the latter's territory. (2) Repeated interference by males in the courtship feeding of another pair. Such behaviour took the form of either physically preventing the transfer of food to the female, or of begging (but not taking food) from the allofeeding male. Similar begging was also directed towards single birds. Because begging females accepted allofed insects, alternative nutritional explanations for their behaviour could not be discounted. Consequently, cases based only on begging evidence by females were excluded. (3) Blocking a pre-breeding pair from access to its nest chamber. In this behaviour the harassing individual would position itself at a nest entrance and either chase or deny access to the occupants as they attempted to enter. (4) Repeated visits, typically accompanied by begging or other vocalizations, to the nest chamber of another pair during the period before egg laying or hatching at the nest of the harasser. (5) In some analyses we also included cases ($N=9$) in which harassment was implied but not directly observed: one member of an incubating pair abandoned its nest and began helping elsewhere while its mate continued to tend the clutch alone.

Received 3 January; accepted 20 February 1992.

- Jacobs, J. A. *Nature* **172**, 297-298 (1953).
- Loper, D. E. *Adv. Geophys.* **26**, 1-34 (1984).
- Jacobs, J. A. *The Earth's Core* (Academic, London, 1987).
- Gubbins, D. *J. Geophys.* **43**, 453-464 (1977).
- Loper, D. E. *Geophys. J. R. astr. Soc.* **54**, 389-404 (1978).
- Gubbins, D., Masters, T. G. & Jacobs, J. A. *Geophys. J. R. astr. Soc.* **59**, 57-99 (1979).
- Loper, D. E. *J. Geomag. Geoelectr.* **43**, 79-91 (1991).
- Verhoogen, J. *Geophys. J. R. astr. Soc.* **4**, 276-281 (1961).
- Braginsky, S. I. *Dokl. Akad. Nauk. SSSR* **149**, 8-10 (1963).
- Stevenson, D., Spohn, T. & Schubert, G. *Icarus* **54**, 466-489 (1983).
- Mollett, S. *Geophys. J. R. astr. Soc.* **76**, 653-666 (1984).
- Stacey, F. D. & Loper, D. E. *Phys. Earth planet. Inter.* **36**, 99-115 (1984).
- Ashby, M. F. & Verrall, R. A. *Phil. Trans. R. Soc. A* **288**, 59-87 (1977).
- O'Connell, R. J. *Tectonophysics* **38**, 119-136 (1977).
- Wu, P. & Peltier, W. R. *Geophys. J. R. astr. Soc.* **70**, 435-486 (1982).
- Hager, B. H. *J. geophys. Res.* **89**, 6003-6015 (1984).
- Gwinn, C. R., Herring, T. A. & Shapiro, I. I. *J. geophys. Res.* **91**, 4755-4765 (1986).
- Lewis, J. S. *Earth planet. Sci. Lett.* **11**, 130-134 (1971).
- Stacey, F. D. *Physics of the Earth* 2nd edn (Wiley, New York, 1977).
- Hale, C. J. *Nature* **329**, 233-237 (1987).
- Masters, T. G. & Shearer, P. M. *J. geophys. Res.* **95**, 691-721 (1990).
- Dziewonski, A. M. & Anderson, D. L. *Phys. Earth planet. Inter.* **25**, 297-356 (1981).
- Stacey, F. D. *Phys. Earth planet. Inter.* **15**, 341-348 (1977).
- Verhoogen, J. *Energetics of the Earth* (National Academy of Science, Washington, DC, 1980).

ACKNOWLEDGEMENTS. We thank B. A. Bolt, D. O. Gough, D. Gubbins, J. A. Jacobs, R. Jeanloz, D. E. Loper, D. P. McKenzie, H. K. Moffatt, W. H. Munk and D. J. Stevenson for helpful comments. This work was supported in part by the British Petroleum Venture Research Unit, and by a Visiting Professorship to H.E.H. in the School of Mathematics, University of New South Wales, Australia.

Harassment was often directed towards both members of pre-breeding or breeding pairs, yet only males (the natal members of each pair) became helpers at the nests of harassers. We therefore considered males to be the ultimate targets of the harassment attempts.

We directly observed 47 cases of harassment, of which 16 resulted in recruitment (defined when the harassee joined the breeding effort of the harasser as a helper). Among pairs, harassed individuals were more than twice as likely to become helpers as were individuals for which harassment was not observed; statistically this difference was of borderline significance ($P=0.034$ by simple univariate contingency test, but, $P=0.12$ by logistic regression analysis controlling for relatedness and age of the harassee). When cases of implied harassment (see 5 above) were included, the difference in likelihood of recruitment became highly significant ($P<0.01$; logistic regression).

The social and demographic profile of observed harassers fits that predicted if the function of the behaviour were to obtain the services of helpers (Table 1). All were members of breeding pairs and thus in a position to benefit by gaining a helper. Virtually all (91%) were male, and three-quarters (72%) were older than the individuals they harassed. Dominance is age- and sex-related in most species of avian cooperative breeders^{6,13,14}. Assuming the same to be true of *M. bullockoides*, old males would have the greatest leverage to influence the behaviour of younger individuals. These findings support the hypothesis that harassment is a form of coercion that increases a breeder's chances of obtaining helpers.

Harassers preferentially selected close genetic kin as targets of their harassment ($P<0.01$; chi-square test of null hypothesis of random targeting). Parents were involved in 36–45% of all harassment attempts (omitting and including inferred cases, respectively; see criterion 5), brothers in 16–19%, and other kin (grandparents, uncles and a half-sibling) in 16–17% (Table 1).

The likelihood of harassment leading to successful recruit-

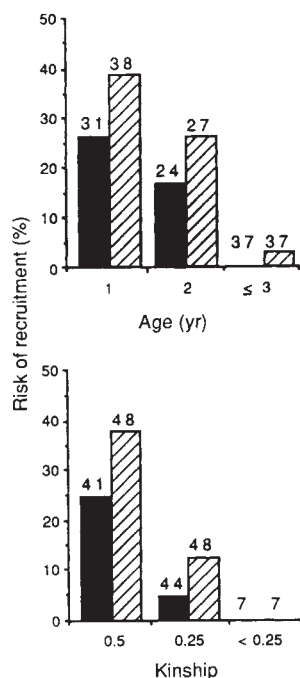


FIG. 1 The effects of age (top) and kinship (bottom) on the probability of paired male bee-eaters being recruited. Kinship is defined as the coefficient of relatedness to the offspring of the harasser pair (the potential recipients of help). Black bars indicate minimum estimate of recruitment risk; hatched bars indicate maximum estimate (see text). Numbers above bars refer to number of potential recruit pairs in each category. Both age and kinship are significant predictors (both $P<0.01$) in a logistic regression analysis of recruitment risk.

ment increased with the difference in age and with the closeness of genetic relatedness between harasser and harassee (Table 1). The highest success rates were achieved by parents with their grown sons (75–84%). Viewed a different way, 69–76% of successful recruitment attempts involved a parent and its offspring, and 62–68% involved father and son.

We estimated the risk of being recruited for different categories of paired birds. Each paired male having an equally aged or older paired relative present in its clan was classified as a potential recruitee. Risk was defined as the per cent of such individuals recruited. In one estimate, only cases based on the five criteria described were included. This presumably underestimates true risk, because some instances of harassment are not observed, and because some uncounted observations of female begging might actually be harassments. A second estimate included all cases where a potential recruitee failed to breed and helped elsewhere in its clan. This presumably overestimates risk, because some pairs may help rather than breed for reasons unrelated to harassment. The true recruitment risk lies somewhere in between.

Figure 1 plots recruitment risk as a function of the age of the potential recruitee and its relatedness to the potential recipients of its help (the offspring of the harassing pair). A yearling paired male (pooling across r values) has a 26–40% probability of having its nesting attempt disrupted. For the subset of such males with breeding parents present in the clan, this risk

TABLE 1 Social and demographic characteristics of harassers

	Number of attempts*	Number successful†	Success rate (%) of attempts‡
Mating status			
Paired	47 (56)	16 (25)	38 (49)
Single	0 (0)	—	—
Age (relative to harassee)			
Older	34 (42)	15 (23)	52 (62)
Same	6 (6)	1 (1)	17 (17)
Younger	7 (8)	0 (0)	0 (0)
Sex§			
Males	43 (51)	15 (23)	39 (50)
Females	4 (5)	1 (2)	25 (40)
Relationship (to harassee) ¶			
Father (+ mother)	13 (20)	9 (16)	75 (84)
Father (+ non-kin)	3 (3)	1 (1)	50 (50)
Mother (+ non-kin)	1 (2)	1 (2)	100 (100)
Brother (+ non-kin)	9 (9)	3 (3)	33 (33)
Grandparent (+ grandparent)	2 (2)	1 (1)	50 (50)
Other kin (+ non-kin)	6 (7)	0 (0)	0 (14)
Nonkin (+ kin)	4 (4)	1 (1)	25 (25)
Nonkin (+ non-kin)	9 (9)	0 (0)	0 (0)
Kinship (to offspring of harasser) ¶			
0.5	13 (20)	9 (16)	75 (84)
0.25	21 (23)	7 (9)	35 (41)
0.125	4 (4)	0 (0)	0 (0)
0.0	9 (9)	0 (0)	0 (0)

First number in each cell represents cases directly observed (criteria 1–4; see text); numbers in parentheses include criterion 5 as well, with harasser defined as the male member of the pair that receives help.

* Includes 16 cases of single harassees.

† Includes 9 cases of single harassees.

‡ Success could not be assigned for five attempts, because the harasser nest failed shortly after the disruption attempt. These cases are omitted in the success rate calculations.

§ These values may overestimate male harassers because of the exclusion of females from category (2) evidence (see text). Cases from category (5) evidence were arbitrarily assigned to male kin if present in the harasser pair.

¶ The identity of the harasser's mate is identified in parentheses.

¶ Kinship assignments assume breeders to be genetic parents of the young they rear.

increases to 42–63%. Risk falls off rapidly as an individual ages and as older close relatives die. By the time an individual is 3 years old, it is virtually immune from harassment, and typically has become a harasser itself (average life expectancy from age of first breeding in this species is 3.7 years).

Continued association between parents and grown offspring (or other close relatives) entails various costs and benefits to both parties^{1–6}. One potential benefit to parents is receipt of help in the rearing of subsequent broods of young. Parents are predicted to encourage such helping whenever they gain more in direct (individual) fitness¹⁵ through the services of the helper than they would in indirect fitness if the potential helper bred independently^{5,9,16}. Among most cooperatively breeding species, opportunities to reproduce independently are severely constrained and both parent and offspring mutually benefit from the continued retention and helping of the latter^{6,7,17–20}. Under certain circumstances, however, fitness payoffs from the two alternative reproductive strategies (helping versus breeding) will be such that the parent benefits, but the offspring loses, by the latter's continued helping (modelled by Emlen⁵). It is under such conditions of genetic conflict that parents are predicted to use their behavioural dominance in an attempt to retain helpers⁵.

White-fronted bee-eaters fulfill these conditions. Breeders benefit greatly from the activities of helpers. The average number of young fledged by unassisted, unrecruited pairs at Nakuru was 0.51; this increased linearly by 0.47 (± 0.06 s.e.) offspring with each helper added²¹. A grown offspring thus contributes far more to its parent's inclusive fitness by helping than if it bred, unaided, on its own (0.47 versus 0.26 offspring equivalents, respectively).

Why don't sons more forcefully resist the recruitment attempts of their parents? The answer presumably lies in the near equity of the inclusive fitness payoffs of the two options available to them. Previous work has demonstrated that the fitness benefit accrued by helpers in this species derives from increasing the production of non-descendant relatives²². By helping rear full sibs, a son will accrue, on average, 0.47 offspring equivalents; by breeding unaided, 0.51. Incorporating probabilistic future fitness gains²³, the values become 0.52 versus 0.58. Thus there is little fitness cost to a son associated with abandoning an early breeding effort and instead helping both parents to rear full sibs.

If this fitness accounting is correct, it predicts that the success of harassment attempts will be much lower when the potential recruits are non-offspring (because the recruit's payoff from helping, in terms of indirect fitness, drops precipitously as relatedness to the nestling recipients decreases). This is confirmed in Table 1.

This same reasoning explains the apparent paradox that the individuals most targeted for harassment are one's closest kin. For more distant relatives (and non-kin), the payoff of continuing to breed will far outweigh that gained through helping. Selection should favour increased resistance to harassment by such individuals (Table 1). Breeders thus will have their greatest behavioural leverage in interactions with their close kin and, in particular, with their own offspring^{5,16}. □

13. Woolfenden, G. E. & Fitzpatrick, J. W. *Condor* **77**, 1–12 (1977).
14. Curry, R. L. *Anim. Behav.* **36**, 1708–1728 (1988).
15. Brown, J. L. & Brown, E. R. in *Natural Selection and Social Behavior: Recent Research and New Theory* (eds Alexander, R. D. & Tinkle, D.) 244–256 (Chiron, New York, 1981).
16. Vehrencamp, S. L. *Anim. Behav.* **31**, 667–682 (1983).
17. Brown, J. L. *Am. Zool.* **14**, 63–80 (1974).
18. Koenig, W. D. & Pitelka, F. A. in *Natural Selection and Social Behavior: Recent Research and New Theory* (eds Alexander, R. D. & Tinkle, D.) 261–280 (Chiron, New York, 1981).
19. Emlen, S. T. *Amer. Nat.* **119**, 29–39 (1982).
20. Woolfenden, G. E. & Fitzpatrick, J. W. *The Florida Scrub Jay: Demography of a Cooperative Breeding Bird* (Princeton Univ. Press, NJ, 1984).
21. Emlen, S. T. & Wrege, P. H. *J. Anim. Ecol.* **60**, 309–326 (1991).
22. Emlen, S. T. & Wrege, P. H. *Behav. Ecol. Sociobiol.* **25**, 303–319 (1989).
23. Mumme, R. L., Koenig, W. D. & Ratnieks, F. L. W. *Anim. Behav.* **38**, 331–343 (1989).

ACKNOWLEDGEMENTS. We thank the government of Kenya and officials at Lake Nakuru National Park for permission to conduct this study; N. Demong, D. Emlen, G. Farley, R. Hegner, E. Kellogg, M. Kinnaird, M. Kipellian, S. Moore, M. Read, G. Tabor and R. Wagner for assistance in the field; K. Loftin and F. Richards for help with data reduction; K. McGowan, R. Mumme, P. Sherman and D. Winkler for comments on the manuscript. This work was supported by the John Simon Guggenheim Foundation, the National Geographic Society and the National Science Foundation.

Clustering of voltage-dependent sodium channels on axons depends on Schwann cell contact

Eun-hye Joe & Kimon Angelides

Departments of Molecular Physiology and Biophysics and Neuroscience, Baylor College of Medicine, One Baylor Plaza, Houston, Texas 77030, USA

IN myelinated nerves, segregation of voltage-dependent sodium channels to nodes of Ranvier is crucial for saltatory conduction along axons^{1–4}. As sodium channels associate⁵ and colocalize with ankyrin at nodes of Ranvier⁶, one possibility is that sodium channels are recruited and immobilized at axonal sites which are specified by the subaxolemmal cytoskeleton, independent of glial cell contact^{7–10}. Alternatively, segregation of channels at distinct sites along the axon may depend on glial cell contact^{11–14}. To resolve this question, we have examined the distribution of sodium channels, ankyrin and spectrin in myelination-competent co-cultures of sensory neurons and Schwann cells by immunofluorescence, using sodium channel-, ankyrin- and spectrin-specific antibodies. In the absence of Schwann cells, sodium channels, ankyrin and spectrin are homogeneously distributed on sensory axons. When Schwann cells are introduced into these cultures, the distribution of sodium channels dramatically changes so that channel clusters on axons are abundant, but ankyrin and spectrin remain homogeneously distributed. Addition of latex beads or Schwann cell membranes does not induce channel clustering. Our results suggest that segregation of sodium channels on axons is highly dependent on interactions with active Schwann cells and that continuing axon-glial interactions are necessary to organize and maintain channel distribution during differentiation of myelinated axons.

We cultured dorsal root ganglion neurons (DRGs) in the presence or absence of Schwann cells^{15–17}. DRGs and accompanying Schwann cells were cultured in a chemically defined medium and were arrested in a basal-lamina-free premyelination stage¹⁵. After plating, associated Schwann cells proliferated, adhered tightly to axons, and made extensive contact with the axonal surface. Formation of basal lamina and myelination were initiated by the addition of serum and ascorbic acid to the medium¹⁵. In the second method, dissociated DRGs were treated with antimitotic reagents to remove associated Schwann cells and fibroblasts, leaving isolated neurons free from Schwann cells. These neurons were later reseeded with pure Schwann cells^{15–18}. Within one week of reseeded, Schwann cells migrated onto axons and made extensive contact but were restrained from forming myelin by withholding ascorbic acid^{15,17}.

The distribution of sodium channels, ankyrin and spectrin were examined under each of these culture conditions by

Received 25 October; accepted 23 December 1991.

1. Hamilton, W. D. *J. Theor. Biol.* **7**, 1–52 (1964).
2. Alexander, R. D. *A. Rev. Ecol. Syst.* **5**, 325–383 (1974).
3. Trivers, R. L. *Am. Zool.* **14**, 249–264 (1974).
4. Vehrencamp, S. L. in *Handbook of Behavioral Neurobiology 3: Social Behavior and Communication* (eds Marler, P. & Vandenbergh, J. G.) 351–394 (Plenum, New York, 1979).
5. Emlen, S. T. *Am. Nat.* **119**, 40–53 (1982).
6. Brown, J. L. *Helping and Communal Breeding in Birds* (Princeton Univ. Press, NJ, 1987).
7. Emlen, S. T. in *Behavioural Ecology: An Evolutionary Approach* (eds Krebs, J. R. & Davies, N. B.) 301–337 (Blackwell, Oxford, 1991).
8. Stacey, P. B. & Koenig, W. D. (eds) *Cooperative Breeding in Birds: Long-Term Studies of Ecology and Behaviour* (Cambridge Univ. Press, UK, 1990).
9. Emlen, S. T. & Vehrencamp, S. L. in *Perspectives in Ornithology* (eds Brush, A. H. & Clark, G. A. Jr) 93–120 (Cambridge Univ. Press, UK, 1983).
10. Emlen, S. T. in *Cooperative Breeding in Birds: Long-Term Studies of Ecology and Behaviour* (eds Stacey, P. B. & Koenig, W. D.) 489–526 (Cambridge Univ. Press, UK, 1990).
11. Hegner, R. E. & Emlen, S. T. *Ecology* **76**, 189–222 (1987).
12. Hegner, R. E., Emlen, S. T. & Demong, N. J. *Nature* **298**, 264–266 (1982).

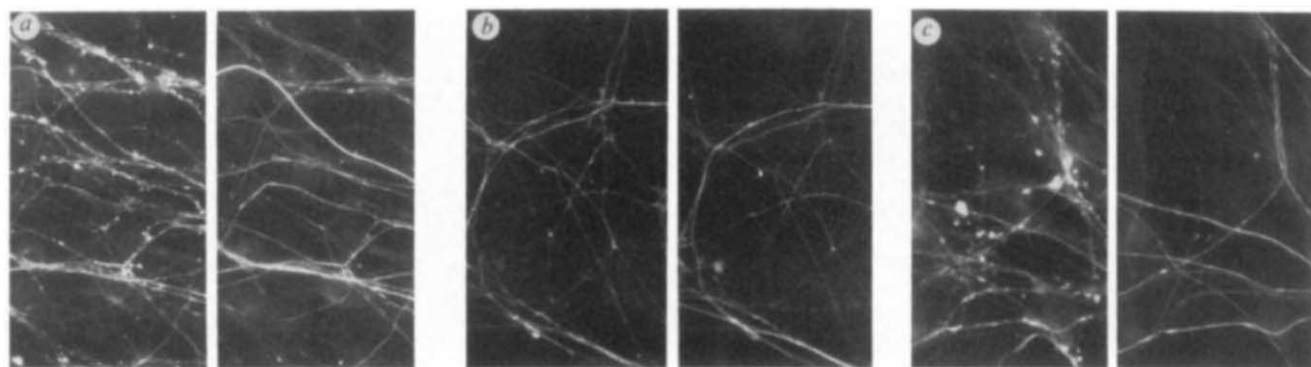
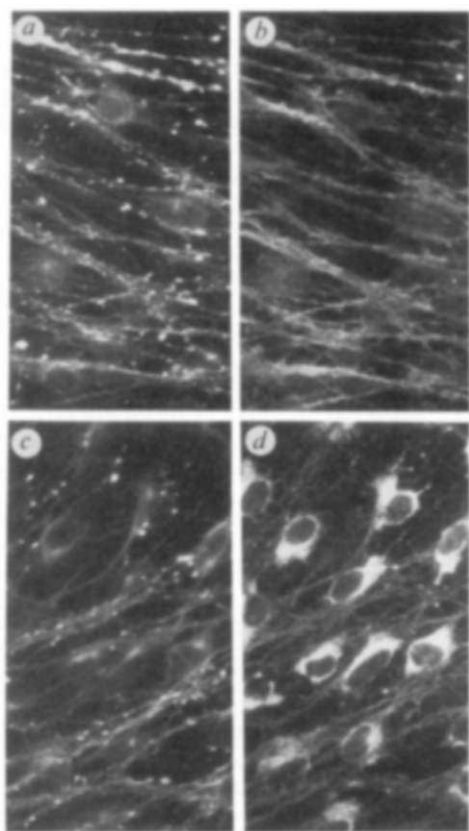


FIG. 1 Distribution of sodium channels (left panels) and ankyrin (right panels) on DRG axons in the presence (a, c) and absence (b) of Schwann cells. a, DRGs accompanied by endogenous Schwann cells (1 week) cultured in chemically defined media. b, Pure DRGs (2 weeks in culture) from which Schwann cells were eliminated by treating with antimitotic reagents. c, Cultures of DRGs one week after re-seeding with purified Schwann cells. In a and c Schwann cell bodies can be seen in the background.

METHODS. DRG sensory neurons were cultured using the method of Eldridge *et al.*¹⁵ with minor modifications. Dorsal root ganglia were removed from 15-day embryonic Sprague-Dawley rats, dissociated with dispase (50 U ml⁻¹, Collaborative Research) and plated on collagen-coated coverslips. DRGs with Schwann cells were cultured in defined media to allow Schwann cells to proliferate. To prepare pure neurons, cultures were treated as described^{15,17} with 10 μ M of 5-fluoro-2'-deoxyuridine. Pure Schwann cells were prepared from rat sciatic nerve according to ref. 18. One week later, Schwann cells were collected and added to pure DRG neuron cultures (3–4 weeks old) and cultured for 1–2 weeks before immunocytochemistry. Polyclonal anti-sodium channel antibody 7493 was prepared and characterized as described¹⁹. The

monoclonal anti-ankyrin antibody F11 was prepared against purified erythrocyte ankyrin and characterized by immunoblotting, immunoprecipitation and immunocytochemistry. Antibody F11 maps to the 72K chymotryptic domain of erythrocyte ankyrin in immunoblots and in immunocytochemistry stains only neurons and their axons. For indirect immunocytochemistry, cells on coverslips were washed with phosphate buffered saline (PBS), fixed with 95% ethanol/5% acetic acid for 30 min at -20°C , washed in PBS, then with 10% goat serum in PBS and rinsed with PBS. Fixation by paraformaldehyde yielded the same distribution. Antibodies were added to cells overnight at room temperature. Sodium channel and ankyrin primary antibodies were labelled with rhodamine and fluorescein conjugate secondary antibodies, respectively. Cells were examined through a 63×1.3 NA objective on a Zeiss Axiovert microscope, and images were captured in digital format through a 1024×1024 CCD camera interfaced to a Perceptics image processor. Sodium channel, ankyrin, and spectrin images were displayed in split screen formats or overlayed in Red/Green/Blue format to facilitate comparison. Images were stored on optical disk and micrographs were printed on a Sony colour video printer.



immunofluorescence microscopy using monoclonal and polyclonal antibodies specific for the sodium channel¹⁹, ankyrin and spectrin. Figure 1 shows the distribution of channels and ankyrin on DRG axons in the presence (Fig. 1 a and c) and absence (Fig. 1 b) of Schwann cells. When neurons are accompanied by endogenous Schwann cells, numerous channel clusters are found on axons (Fig. 1a, left), but ankyrin (Fig. 1a, right) and spectrin (not shown) are uniformly distributed.

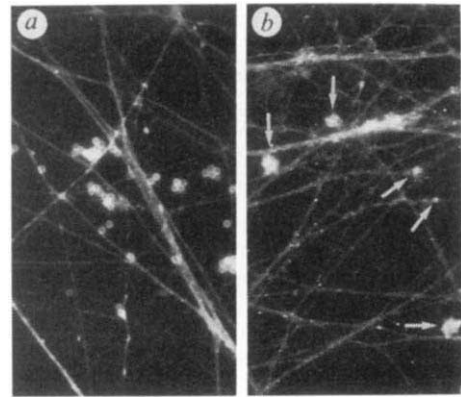
We next tested whether sodium channels and ankyrin had similar distributions in cultures of pure DRG neurons from which Schwann cells had been eliminated. In contrast to the channel clusters seen in neuron-Schwann cell co-cultures, in the absence of Schwann cells, sodium channels (Fig. 1b, left), ankyrin (Fig. 1b, right) and spectrin (not shown) are homogeneously distributed along the axon. When we introduced pure Schwann cells back into these neuron cultures, within one week the channel staining pattern changed dramatically from homogeneous to clustered (Fig. 1c, left). On every axon examined, the distribution of ankyrin (Fig. 1c, right) and spectrin

◀ FIG. 2 Distribution of sodium channels (a, c), NCAM (b), and sCon A receptors (d) on DRG-Schwann cell co-cultures.

METHODS. Cells were labelled with a polyclonal anti-sodium channel antibody followed by a rhodamine-conjugated secondary antibody, and a monoclonal antibody against NCAM (1:100 dilution) followed by a fluorescein-conjugated secondary antibody. For sCon A labelling, 30 μ g ml⁻¹ of rhodamine succinylated concanavalin-A (Vector Laboratories, Inc.) was added to cultures for 10 min at 4°C , washed and then fixed with 95% ethanol/5% acetic acid. After fixation, cells were labelled with polyclonal anti-sodium channel antibody as in Fig. 1, followed by labelling with fluorescein-conjugated secondary antibodies. Although sCon A was applied to cells at 4°C , the internalization of some sCon A receptors in Schwann cell bodies could not be inhibited.

FIG. 3 Distribution of sodium channels on DRG axons in the presence of latex beads (a) and Schwann cell membranes (b). Schwann cell membranes are denoted by arrows.

METHODS. Rhodamine-conjugated latex beads (2.7 μm , Polysciences, Inc.) were added to 2-week-old pure sensory neurons. Schwann cell membranes were prepared from 1-week-old Schwann cells. Schwann cells were harvested from cultures and triturated in 30% Hank's balanced salt solution with a 25 gauge needle. Disrupted Schwann cells were applied to a 0.25 M sucrose pad and centrifuged at 2,000g for 5 min. Crude membranes were recovered from the supernatant and were added to isolated neurons (2 weeks in culture). One week after addition of latex beads and Schwann cell membranes, cells were fixed and labelled with anti-sodium channel antibodies (fluorescein-conjugated secondary antibody). In the images presented the latex bead image is overlaid on the sodium channel image.



(not shown) continued to be homogeneous. During the first few days after reseeding very few axons exhibited punctate channel staining. As Schwann cells proliferated to areas of the culture occupied by axons, the density of channel clusters increased. Analysis of many images revealed that the density of channel clusters on axons was related to the Schwann cell density. In regions densely populated by Schwann cells we found ~ 405 sodium channel clusters per mm^2 , while in regions of the culture where Schwann cells were sparse there were < 10 clusters per mm^2 .

We next asked whether the clustering of sodium channels induced by Schwann cells showed any molecular specificity for the sodium channel or was a general feature of neuronal membrane glycoproteins. Double-labelling of cultures for sodium channels and NCAM, a glycoprotein that is concentrated at nodes²⁰ or succinylated concanavalin A receptors (Fig. 2) showed that while sodium channels were clustered (Fig. 2a and c), both NCAM (Fig. 2b) and succinylated concanavalin A receptors (Fig. 2d) have distinctly different distributions. To test whether channel clustering arose from contact itself or from contact with a Schwann cell membrane component, we added latex beads (Fig. 3a) and Schwann cell membranes (Fig. 3b) to pure neurons. In both of these treatments, the homogeneous distribution of sodium channels and of ankyrin persisted.

The extent to which segregation of sodium channels to the node of Ranvier depends on the Schwann cell is an important issue in developmental neurobiology. Previous studies on the distribution of intramembrane particles and heavy metal stains^{7,9,21} have suggested that the differentiation of the axon into nodal and internodal regions occurs early in axon development. Clusters of intramembrane particles^{7,9,10,22}, Fe-FeCN staining²¹, and Na^+ current²³ have been observed in developing axons. When myelination is delayed²⁴, maturation of the axon continues with the establishment of a regular nodal geometry, increased axolemmal undercoating and changes in conduction properties, features which imply that segregation of sodium channels in axons develops independently of glial cells. But other studies have suggested that when axon-Schwann cell relationships are disrupted, intramembrane particles or intense Fe-FeCN staining at nodes is lost^{25,26} suggesting that continuing axon-glial interactions are required to develop and maintain high sodium channel density on the axon. The work presented here using sodium channel-specific antibodies shows that contact between Schwann cell and axon is essential in regulating sodium channel distribution on the axon membrane and that the neuron does not sequester sodium channels at specific axonal sites independent of the Schwann cell.

We reported previously that ankyrin associates with, and links sodium channels to the underlying cytoskeleton⁵, and proposed that this linkage could help in the development and control of channel distribution on the axon^{5,27}. The idea that sodium channels are restrained by cytoskeletal linkages is compatible

with studies in rat sciatic nerve which have shown that an isoform of ankyrin is concentrated at nodes of Ranvier⁶. Even though our anti-ankyrin antibody recognizes an ankyrin isoform that associates with sodium channels and is localized at mature nodes in these cultures, this distribution is established only during the terminal stages of myelination.

There is molecular specificity in the mechanisms by which Schwann cells regulate protein topography over the axon membrane. It is not known how Schwann cells modulate sodium channel clustering on the axon without affecting other axonal membrane proteins, but the process must involve specific molecular recognition between the Schwann cell and axon membrane. It is possible that the sodium channel clusters arise by a specific suppression of channel density in regions ensheathed by Schwann cells²⁸. Alternatively, it is known that components in the axon membrane are mitogenic for Schwann cells²⁹ and that the Schwann cell undergoes a programmed expression of specific adhesion molecules essential for myelination when in contact with axons³⁰⁻³². Elements contributed by the Schwann cell and/or extracellular matrix, similar to those that induce acetylcholine receptor clustering at the neuromuscular junction³³, could promote initial clusters of sodium channels as the axon membrane develops. □

Received 11 October 1991; accepted 3 January 1992.

1. Waxman, S. G. & Ritchie, J. M. *Science* **228**, 1502-1507 (1985).
2. Black, J. A., Kocsis, J. D. & Waxman, S. G. *Trends Neurosci.* **13**, 48-54 (1990).
3. Bostock, H. & Sears, T. A. *J. Physiol., Lond.* **280**, 273-301 (1978).
4. Rasminsky, M. & Sears, T. A. *J. Physiol., Lond.* **227**, 323-349 (1972).
5. Srinivasan, Y., Elmer, L., Davis, J., Bennett, V. & Angelides, K. *J. Nature* **333**, 177-180 (1988).
6. Kordeli, E., Davis, J., Trapp, B. & Bennett, V. *J. Cell Biol.* **110**, 1341-1352 (1990).
7. Ellisman, M. H. *J. Neurocytol.* **8**, 719-748 (1979).
8. LeBeau, J. M., Powell, H. C. & Ellisman, M. H. *J. Neurocytol.* **16**, 347-358 (1987).
9. Wiley-Livinston, C. A. & Ellisman, M. H. *Dev Biol.* **79**, 334-355 (1980).
10. Wiley, C. A. & Ellisman, M. H. *J. Cell Biol.* **84**, 261-280 (1980).
11. Rosenbluth, J. *J. Neurocytol.* **8**, 655-672 (1979).
12. Rosenbluth, J. *Int. J. devl Neurosci.* **6**, 3-24 (1988).
13. Black, J. A., Waxman, S. G. & Hildebrand, C. *J. Neurocytol.* **14**, 887-909 (1985).
14. Bigbee, J. W. & Foster, R. E. *Brain Res.* **494**, 182-186 (1989).
15. Eldridge, C. F., Bunge, M. B., Bunge, R. P. & Wood, P. M. *J. Cell Biol.* **105**, 1023-1034 (1987).
16. Eldridge, C. F., Bunge, M. B. & Bunge, R. P. *J. Neurosci.* **9**, 625-638 (1989).
17. Clark, M. B. & Bunge, M. B. *Dev Biol.* **133**, 393-404 (1989).
18. Brookes, J. P., Fields, K. P. & Raff, M. C. *Brain Res.* **165**, 105-108 (1979).
19. Elmer, L. W., Black, J. A., Waxman, S. G. & Angelides, K. *J. Brain Res.* **532**, 222-231 (1990).
20. Reiger, F., Daniloff, J. K., Pincon-Raymond, M., Crosin, K. L., Grumet, M. & Edelman, G. M. *J. Cell Biol.* **103**, 379-391 (1986).
21. Waxman, S. G. & Foster, R. E. *Proc. R. Soc. B209*, 441-446 (1980).
22. Black, J. A., Foster, R. E. & Waxman, S. G. *J. Neurocytol.* **10**, 981-993 (1981).
23. Smith, K. J., Bostock, H., and Hall, S. M. *J. Neurol. Sci.* **54**, 3-31, 1982.
24. Black, J. A., Sims, T. J., Waxman, S. G. & Gilmore, S. A. *J. Neurocytol.* **14**, 79-104 (1985).
25. Foster, R. E., Whalen, C. C. & Waxman, S. G. *Science* **210**, 661-663 (1980).
26. Rosenbluth, J. & Blakemore, W. *Neurosci. Lett.* **48**, 171-177 (1984).
27. Angelides, K. J., Loftus, D., Elmer, L. W. & Elson, E. L. *J. Cell Biol.* **106**, 1911-1925 (1988).
28. Black, J. A., Waxman, S. G., Sims, T. J. & Gilmore, S. A. *J. Neurocytol.* **15**, 745-761 (1986).
29. Ratner, N., Glaser, L. & Bunge, R. P. *J. Cell Biol.* **98**, 1150-1155 (1984).
30. Ranscht, B., Wood, P. M. & Bunge, R. P. *J. Neurosci.* **7**, 2936-2947 (1987).
31. Owens, G. & Bunge, R. P. *Glia* **2**, 119-128 (1989).
32. Wood, P. M., Schachner, M. & Bunge, R. P. *J. Neurosci.* **10**, 3635-3645 (1990).
33. Froehner, S. C. *J. Cell Biol.* **113**, 1133-1144 (1991).

ACKNOWLEDGEMENTS. We thank B. Hicks for discussions. Supported by the National Multiple Sclerosis Society and the NIH.

Rhodopsin inactivation is a modulated process in *Limulus* photoreceptors

Edwin A. Richard & John E. Lisman

Department of Biology, Brandeis University, Waltham, Massachusetts 02254, USA

MANY G-protein-coupled receptors are only transiently active because an inactivation process stops the receptor from activating G protein molecules¹⁻³. Although this inactivation has been investigated *in vitro*, the real kinetics of the process can only be obtained from intact cells. Here we describe a method for measuring the inactivation of rhodopsin in intact photoreceptors and the application of this method to the ultraviolet rhodopsin of *Limulus* median eye⁴. The results show that the inactivation process is very rapid (<150 ms) and occurs well before the peak of the receptor potential. We have also investigated whether the inactivation process can itself be modulated. Our results show that light-adaptation accelerates inactivation by about 10-fold, providing evidence that G-protein-mediated transduction can be modulated at this first stage.

Rhodopsin is excited by the isomerization of the 11-*cis* chromophore to the all-*trans* state. This is followed by a rapid molecular transition to the metarhodopsin state, which activates G proteins until metarhodopsin is biochemically inactivated^{1,2,5-10}. Our experimental method takes advantage of the fact that metarhodopsin can be prematurely inactivated by wavelengths of light that convert the all-*trans* chromophore back to 11-*cis* and thereby photoregenerate rhodopsin⁷. If such photochemical inactivation is triggered before normal inactivation, the duration of the active state will be shortened. This will reduce the number of G proteins activated and reduce the size of the receptor potential. *Limulus* ultraviolet receptors are particularly suitable for this approach because of the wide spectral separation of rhodopsin ($\lambda_{\max} = 360$ nm) and metarhodopsin ($\lambda_{\max} = 470$ nm)⁸. This makes it possible to excite rhodopsin selectively using ultraviolet light and to photoregenerate rhodopsin selectively using long-wavelength light. Results indicating that the duration of metarhodopsin activity can be shortened by photoregeneration are shown in Fig. 1a. A dim ultraviolet flash was used to convert a small fraction of the rhodopsin to metarhodopsin. By itself, this stimulus (ultraviolet light alone) generates an 11-mV receptor potential. If the ultraviolet flash is followed 10 ms later by an intense, photoregenerating flash, the receptor potential is greatly reduced, indicating that the duration of metarhodopsin activity is shortened. As the interval between the flashes is increased, the amplitude of the response increases. For example, if the regenerating flash is given 60 ms after the ultraviolet flash (Fig. 1a), the regenerating flash has only a small effect. This indicates that by this time metarhodopsin must have already been substantially inactivated by the endogenous biochemical process.

To quantify the time course of inactivation, the area (integral) of the response to the ultraviolet flash was measured as a function of the delay of the photoregenerating flash. For small responses, the area is a linear (Fig. 1b) and time-invariant (Fig. 1c) measure of the photoresponse. The dependence of the photoresponse on the delay to the photoregenerating flash was fitted to a function having two exponential time constants, τ_1 and τ_2 (see Fig. 2 legend). The time constant τ_1 describes the rapid formation of metarhodopsin from a short-wavelength-absorbing pigment intermediate¹¹⁻¹³. This time constant (6.2 ms) was measured from the leading edge of the early receptor potential⁸ (Fig. 2b). The apparent time constant τ_2 for the natural process of receptor inactivation is determined by the rising part of the curve in Fig. 2a. In this experiment, τ_2 was 35 ms. Comparison of the time

course of inactivation to the time course of the receptor potential itself (Fig. 2c) shows that inactivation is complete before the peak of the receptor potential. Indeed the active state of the pigment occurs primarily during the latent period of the receptor potential. The lifetime of active metarhodopsin in *Limulus* is considerably briefer than the value of 1.7 s measured in vertebrate rods using a less direct assay¹⁴. This difference is consistent with the much shorter overall timescale of the response in invertebrates.

Because increasing the rate of inactivation would lower the number of G proteins activated, it would seem likely *a priori* that light-adaptation, which reduces the overall gain of transduction, might increase the rate of inactivation. Figure 3 shows that

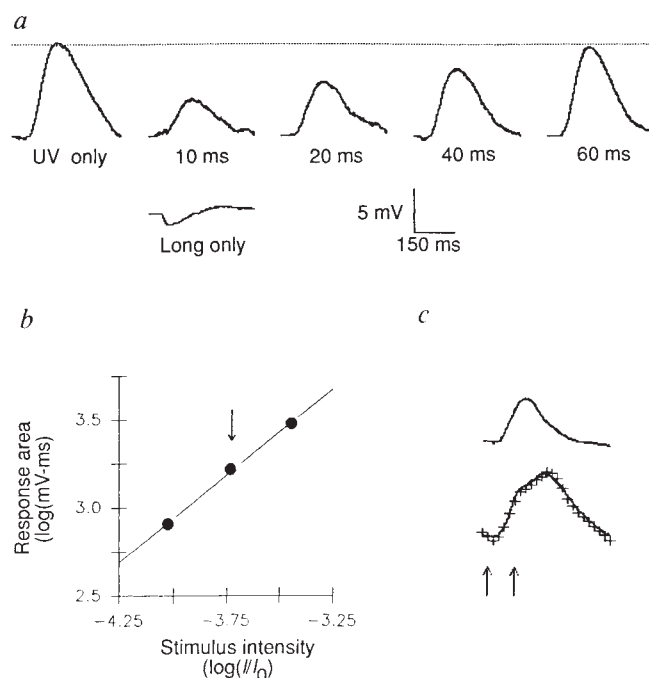


FIG. 1 The response to an ultraviolet (UV) flash can be reduced by a photoregenerating flash. **a**, The first trace is the response to a UV flash given at the beginning of the trace (UV only). Subsequent traces are the response to a UV flash of the same intensity followed by a regenerating flash given with the indicated delay. These traces were corrected for the small response generated by the regenerating flash alone (Long only; $N=8$). In controls, a photoregenerating flash was given just before a UV flash; the response was consistent with linear superposition of the responses to flashes given separately. **b**, Photoresponse area was linear with flash intensity. Data points are means ($N=3-8$); standard error bars were less than the symbol size (slope, 0.98). The arrow indicates the flash intensity used in **a** and Fig. 2. **c**, There were no time-dependent changes in responsiveness. The top trace shows the averaged photoresponse to a single flash (at first arrow; $N=3$); the lower trace shows the averaged response to two sequential flashes (at arrows; $N=3$) superposed with the sum of two averaged responses predicted from the first flash, assuming linear superposition and no change in responsiveness (crosses). The entire experiment was done in the presence of an adapting light (4.3×10^4 effective photons absorbed per second) that desensitized the cell by 0.7 log units. **METHODS**. Cells were impaled at room temperature (20–23 °C) with a single electrode containing 2 M potassium acetate as described⁸. The UV stimulus (at 337 nm), which always excited less than 10^{-4} rhodopsin molecules, was a pulse from a VSL 337 N₂ laser (Laser Science). The output, I_0 was attenuated with calibrated neutral density filters. The regenerating flash source was a Strobex 278 Xe flash lamp (Chadwick-Helmuth) filtered to >495 nm with multiple Schott cut-on filters. The regenerating flash itself activated <1,000 rhodopsin molecules which, in dark-adapted cells, leads to a receptor potential that cannot be separated from the response to the UV flash. To minimize the depolarization produced by the regenerating flash, experiments were done in the presence of a moderate-intensity UV-adapting light. The rate of effective photon absorptions was extrapolated from lower light intensities where the rate of single photon responses could be directly measured.

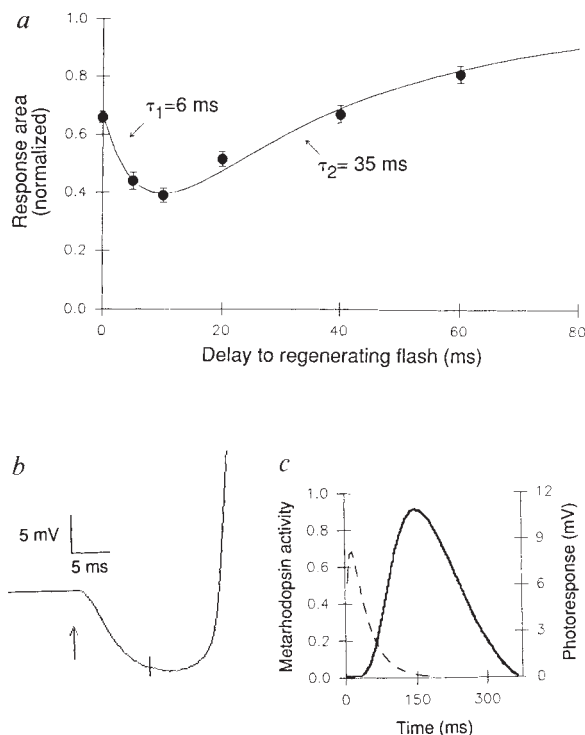


FIG. 2 Time course of metarhodopsin activity. *a*, Photoresponse area was dependent on the delay of the photoregenerating flash. Data points are means with standard errors ($N=8$). The solid line is a least-squares curve fit to

$$1 + A \exp(-t/\tau_1) - B \exp(-t/\tau_2) \quad (1)$$

Same cell and adapting light intensity as used for Fig. 1. *b*, The exponential time constant τ_1 was measured from the leading edge of the negative early receptor potential, a short latency signal generated directly by conformational changes in the pigment molecule (stimulus at arrow)²⁴; τ_1 was corrected for the membrane time constant, as measured by the charging curve evoked by a hyperpolarizing current pulse. *c*, Metarhodopsin activity (dashed line) ended early in the photoresponse (solid line). Fraction of active metarhodopsin molecules calculated relative to the total number of photoisomerized molecules from the model below (same conditions as in *a*).

METHODS. In equation (1) A and B are functions of τ_1 , τ_2 , the duration of the regenerating flash, and the fractions of intermediate and metarhodopsin that are converted back to rhodopsin by the regenerating flash. The 2-ms duration of the regenerating flash means that some metarhodopsin will form from the intermediate and be converted back to rhodopsin even when the regenerating flash is given at $t=0$. The data was best fitted by assuming that the regenerating flash converted a small fraction of the intermediate to rhodopsin. The regenerating flash converted 84% of metarhodopsin to rhodopsin, as calculated from the amplitude of consecutive positive early receptor potentials⁸. The curve fit in *a* is based on the assumption that the short-lived pigment intermediate has no activity; fits based on the alternative assumption yielded very similar estimates of τ_2 .

FIG. 3 Light-adaptation accelerated inactivation. Measurements were made at adapting intensities of 4.3×10^4 effective photons absorbed per second ($\circ$) (see Fig. 2*a*), 4.3×10^6 effective photons absorbed per second (∇), and then again at the first intensity ($\square$). Data points ($N=6-8$) and curve fitting were as in Fig. 2.

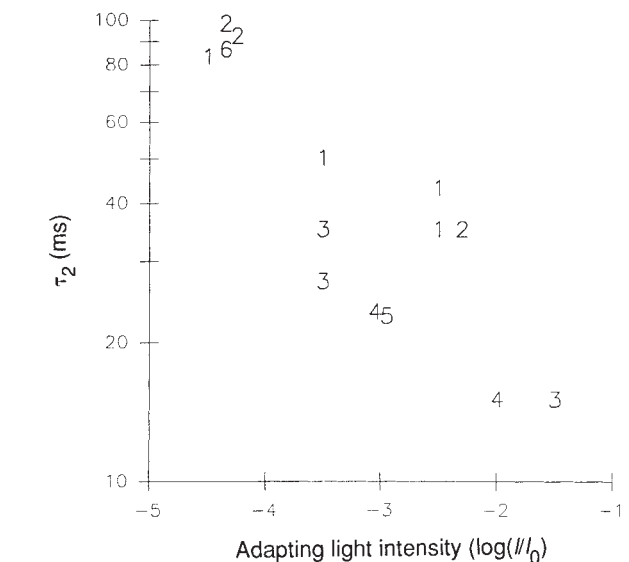
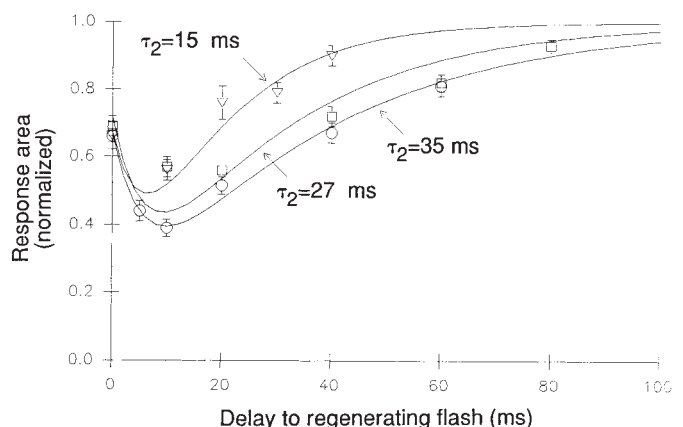


FIG. 4 Active metarhodopsin lifetime as a function of adapting light intensities. Metarhodopsin lifetime was determined as for Fig. 2. Numerals identify data from individual cells. Results from cell 3 were used for the other figures. The number of adapting light intensities investigated in each cell was limited by the 1.5 h required to measure a single τ_2 . This time was determined by the recharging rate of the Xe flash lamp (1 min). Only cells maintaining stable responses over the data acquisition period were used. I_0 was equivalent to 1.4×10^3 effective photons absorbed per second.

this is the case. Increasing the intensity of a steady adapting light caused a reversible decrease in the lifetime of metarhodopsin activity. The lifetime was measured over a range of adapting light intensity (14 measurements in 6 cells) (Fig. 4). The four cells from which multiple measurements were made gave a 2.3 (± 0.3)-fold decrease (mean $\pm$ s.d.) in lifetime per 100-fold increase in intensity. Overall, the inactivation process is modulated over about a 10-fold range (Fig. 4). The actual range could be larger as measurements were not made in the dark-adapted state for reasons discussed in the legend to Fig. 1. These results suggest that the large gain reductions that occur during light-adaptation may be due in part to a reduction in the number of G proteins activated by each excited rhodopsin.

The mechanisms of inactivation and its modulation remain to be determined. Work on vertebrate rhodopsin and β -adrenergic receptors indicates that the inactivation process involves phosphorylation of the receptor by a specialized kinase¹⁵⁻¹⁷ followed by the binding of arrestin^{5,6,18}, which is also present in invertebrate photoreceptors¹⁹⁻²³. The light-dependent phosphorylation of invertebrate arrestin^{22,23} is a plausible candidate for the process modulating inactivation.

Our measurements have given direct information about the



timing of inactivation in living cells and revealed a new process, the modulation of receptor inactivation. This modulation may

be an important way of regulating the gain of G-protein-mediated transduction. □

Received 3 October 1991; accepted 17 January 1992.

1. Lieberman, P. A. & Pugh, E. N. Jr *Nature* **287**, 734–736 (1980).
2. Sitarumayya, A. & Lieberman, P. A. *J. biol. Chem.* **258**, 1205–1209 (1983).
3. Sibley, D. R., Strasser, R. H., Caron, M. G. & Lefkowitz, R. J. *J. biol. Chem.* **260**, 3883–3886 (1985).
4. Nolte, J. & Brown, J. E. *J. gen. Physiol.* **59**, 167–185 (1972).
5. Kühn, H., Hall, S. W. & Wilden, U. *FEBS Lett.* **176**, 473–478 (1984).
6. Wilden, U., Hall, S. W. & Kühn, H. *Proc. natn. Acad. Sci. U.S.A.* **83**, 1174–1178 (1986).
7. Hamdorf, K. & Kirschfeld, K. *Nature* **283**, 859–860 (1980).
8. Lisman, J. *J. gen. Physiol.* **85**, 171–187 (1985).
9. Levine, E., Crain, E., Robinson, P. & Lisman, J. *J. gen. Physiol.* **90**, 575–586 (1987).
10. Kahana, A., Robinson, P. R. & Lisman, J. E. *Biol. Bull.* **179a**, 230 (1990).
11. Azuma, K., Azuma, M. & Suzuki, T. *Biochim. biophys. Acta* **393**, 520–530 (1975).
12. Tokunaga, F., Schichida, Y. & Yoshizawa, T. *FEBS Lett.* **55**, 229–232 (1975).
13. Kirschfeld, K., Feiler, R. & Minke, B. *Z. Naturforschung* **33c**, 1009–1010 (1978).

14. Pepperberg, D. R. et al. *Visual Neurosci.* **8**, 9–18 (1992).
15. Bownds, D., Dawes, J., Miller, J. & Stahman, M. *Nature New Biol.* **237**, 125–127 (1972).
16. Kühn, H. *Nature* **250**, 588–590 (1974).
17. Benovic, J. L., Strasser, R. H., Caron, M. G. & Lefkowitz, R. J. *Proc. natn. Acad. Sci. U.S.A.* **83**, 2797–2801 (1986).
18. Lohse, M. J., Benovic, J. L., Codina, J., Caron, M. G. & Lefkowitz, R. J. *Science* **248**, 1547–1550 (1990).
19. Paulsen, R. & Bentrop, P. *J. comp. Physiol. A* **155**, 39–45 (1984).
20. Smith, D. P., Shieh, B.-H. & Zuker, C. S. *Proc. natn. Acad. Sci. U.S.A.* **87**, 1003–1007 (1990).
21. Hyde, D. R. et al. *Proc. natn. Acad. Sci. U.S.A.* **87**, 1008–1012 (1990).
22. Yamada, T. et al. *Science* **248**, 483–486 (1990).
23. LeVine, H. et al. *Mech. Dev.* **33**, 19–26 (1991).
24. Hagins, W. A. & McLaughly, R. E. *Science* **157**, 813–816 (1967).

ACKNOWLEDGEMENTS. We thank M. Erickson, G. Fain and E. Marder for reading the manuscript. This work was supported by the NIH. E.A.R. was supported by a training grant from the NIH.

Susceptibility of β_2 -microglobulin-deficient mice to *Trypanosoma cruzi* infection

R. L. Tarleton*, B. H. Koller†, A. Latour† & M. Postan*

* Department of Zoology, University of Georgia, Athens, Georgia 30602, USA

† Department of Pathology, University of North Carolina, Chapel Hill, North Carolina 27599, USA

THE β_2 -microglobulin (β_2m) protein associates with the products of the class I major histocompatibility (MHC) loci; this combination functions in the thymic development of and antigen presentation to CD8⁺ T cells. Mice in which the β_2m gene has been disrupted by homologous recombination fail to express class I MHC gene products, and therefore lack CD8⁺ T cells and measurable cytotoxic T-cell responses^{1,2}. However, β_2m ^{−/−} mice appear to have normal development of both CD4⁺ α/β T-cell receptor (TCR⁺) and γ/δ TCR⁺ T cells and are not overtly more susceptible than β_2m ^{+/+} mice to potential environmental agents of infection^{1,2} or to experimental viral infection³. Here we show that β_2m ^{−/−} mice suffer high parasitaemias and early death when infected with the obligate cytoplasmic protozoan parasite *Trypanosoma cruzi*. Despite this increased susceptibility, the β_2m ^{−/−} mice are more respon-

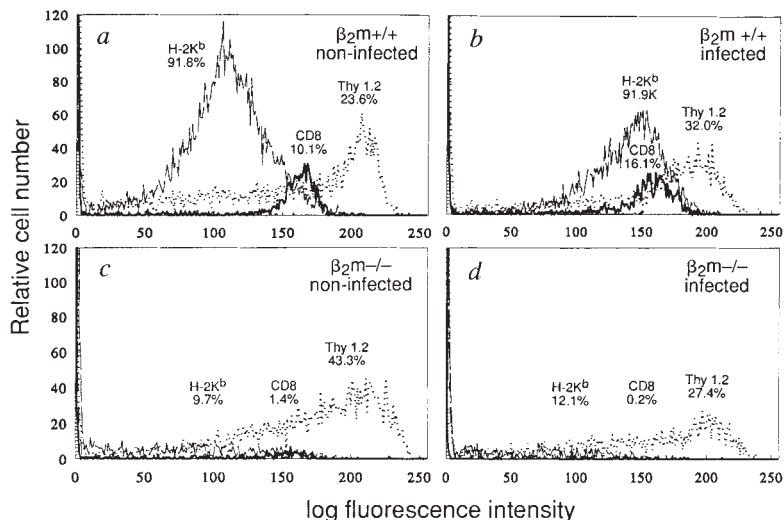
sive than their β_2m ^{+/+} littermates in terms of lymphokine production, making higher levels of both interleukin-2 and interferon- γ in response to mitogen stimulation. In addition, the β_2m ^{−/−} mice show essentially no inflammatory response in parasite-infected tissues. These results confirm previous experiments on mice depleted of CD8⁺ cells using antibody treatment⁴ in demonstrating the importance of CD8⁺ T cells in immune protection in *T. cruzi* infection. They also implicate CD8⁺ T cells and/or class I MHC molecules in regulation of lymphokine production and recruitment of inflammatory cells.

The absence of class I MHC and CD8⁺ T cells was confirmed in both *T. cruzi*-infected and non-infected homozygous β_2m -deficient mice by flow cytometric analysis (Fig. 1). Although *T. cruzi*-infected mice exhibit increased class I expression (mean fluorescence intensity of 108 for uninfected, and 137 for infected mice, perhaps a result of high *in vivo* levels of interferon- γ (IFN- γ ; ref. 5), *T. cruzi*-infected β_2m -deficient mice fail to express class I antigens and lack mature CD8⁺ T cells. Parasitaemias are significantly higher in β_2m ^{−/−} mice than in $+/+$ and $+/-$ controls, and all β_2m -deficient mice succumb to the infection in the acute phase whereas β_2m -competent mice survive (Fig. 2). We have repeated these mortality experiments using a total of over 40 $+/+$, $+/-$ and $-/-$ mice and have obtained nearly identical results. These results confirm our previous studies documenting the requirement for a functional CD8⁺ T-cell compartment in the mouse for survival of *T. cruzi* infection⁴.

In addition to serving as effectors in the immune response to

FIG. 1 Flow cytometric analysis of H-2K^b, Thy1, and CD8 on spleen cells from a, β_2m ^{+/+} non-infected mice; b, β_2m ^{+/+}, day 18 of infection; c, β_2m ^{−/−} non-infected mice; d, β_2m ^{−/−}, day 18 of infection.

METHODS. The mice used in this study were derived from chimaeric mice in which the β_2m gene had been disrupted by homologous recombination¹. Mice were infected by the intraperitoneal route with 10³ blood-form trypomastigotes of the Brazil strain of *T. cruzi* obtained from infected C3H/HeSnJ donors. On day 18 post-infection, infected and non-infected β_2m ^{+/+}, $+/−$ or $−/−$ mice were killed and their spleens processed to obtain a single-cell suspension. Nucleated spleen cells (10⁶) were pelleted in 1.5-ml conical test tubes and washed once in staining medium (PBS with 1% BSA and 0.1% sodium azide) at 4 °C. Cells were then resuspended in 50 μ l of either fluorescein isothiocyanate-labelled (FITC) rat anti-mouse anti-CD8 (H35.17.2), FITC-labelled rat anti-mouse Thy 1.2 (30-H12), or FITC-labelled rat anti-mouse H-2K^b (PharMingen) diluted in staining medium and incubated on ice for 30 min. Following incubation, cells were washed once with staining medium before analysis. For each sample, 20,000 cells were analysed by flow cytometry on an EPICS 753 (Coulter Electronics). Results from 4 individual mice of 12 analysed are shown. The relative cell number indicates the percentage of cells expressing each marker. Background staining was assessed using isotype matched FITC-labelled rat immunoglobulin (PharMingen) and is shown



as a dashed line. The background staining with this antibody was <2% for all samples.

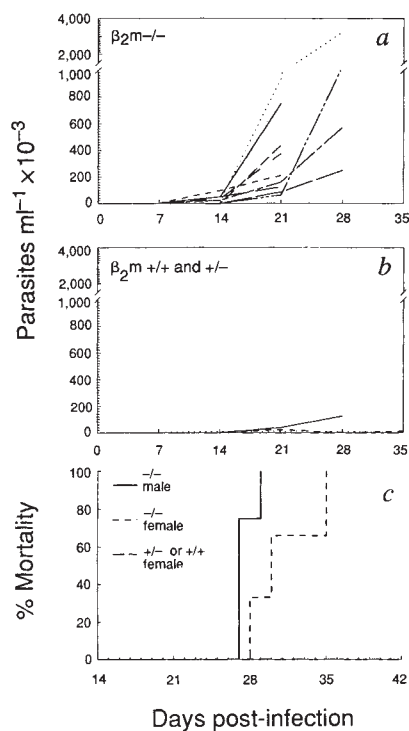


FIG. 2 Parasitaemia (a and b) and mortality (c) of $\beta_2m^{-/-}$, $+/+$ and $+/-$ mice infected with *T. cruzi*.

METHODS. Mice infected as described in Fig. 1 were housed in microisolator cages (Lab Products) and checked daily for deaths. Blood parasite levels were determined weekly in individual mice using haemocytometer counts of 10 μ l tail-vein blood diluted in an ammonium chloride solution. Each line represents the parasitaemia on a single mouse (8 $-/-$ and 5 $+/+$ mice were used in this experiment).

T. cruzi, CD8⁺ T cells have also been implicated in the prominent suppression of immune responses that occurs during the acute phase of infection, particularly in suppression of interleukin-2 (IL-2) production⁶. Our initial experiments indicated that spleen cells from $\beta_2m^{-/-}$ mice infected for 18 days produce significantly higher amounts of IL-2 in response to mitogenic stimulation when compared with $+/+$ and $+/-$ mice (results not shown). This conclusion was confirmed by further analysis of IL-2 production at different times during infection (Fig. 3a).

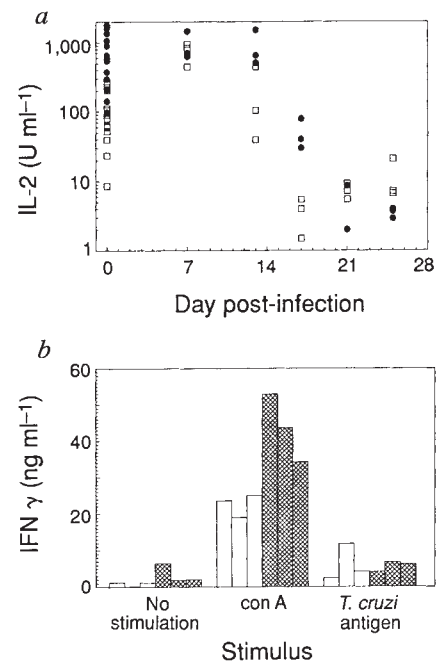


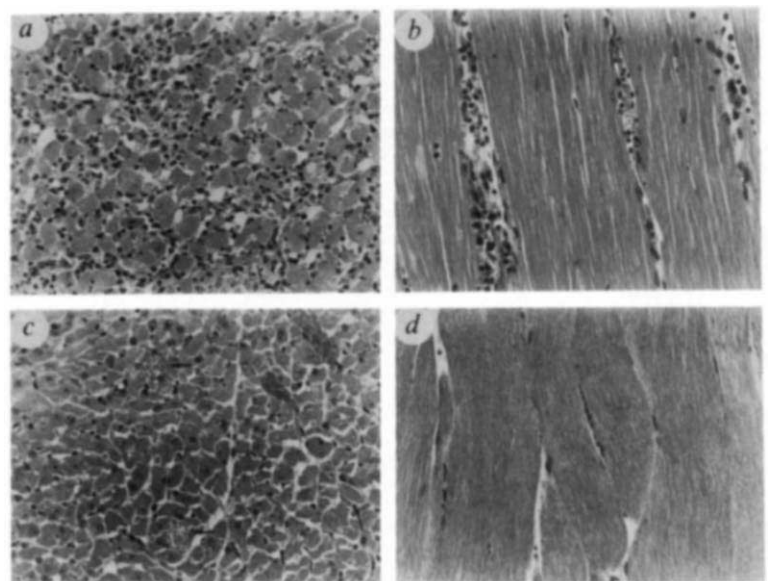
FIG. 3 IL-2 (a) and interferon- γ (b) production by spleen cells from mice infected with *T. cruzi*. a, Squares, $+/+$, $+/-$; filled circles, $-/-$. b, open bars, $+/+$ or $+/-$; hatched bars, $-/-$.

METHODS. Spleen cells (10^7 ml⁻¹) obtained from normal mice or mice infected for various periods of time (a) or for 18 days (b) were cultured without any stimulus or with 5 μ g ml⁻¹ of concavalin A (con A; for IL-2 and IFN- γ) or 10 μ g of a *T. cruzi* trypomastigote sonicate (for IFN- γ). Supernatants from these cultures were collected 48 h after initiation and stored at -20°C . IL-2 levels were measured using the IL-2-dependent HT-2 cell line⁶ and IFN- γ concentrations determined using a sandwich ELISA⁵. Results from individual mice are shown. IL-2 in supernatants from unstimulated cells was below the level of detection (<1 U ml⁻¹).

It is also clear that uninfected $\beta_2m^{-/-}$ mice have a generally higher potential for IL-2 production and eventually become suppressed for IL-2 production like $+/+$ and $+/-$ mice. These results support previous studies demonstrating a role for factors other than CD8⁺ T cells in immunosuppression in *T. cruzi* infection⁷⁻¹⁰. *T. cruzi*-infected β_2m -deficient mice are also better *in vitro* producers of IFN- γ than $+/+$ and $+/-$ mice (Fig. 3b), exhibit spontaneous IFN- γ production *in vitro* (Fig. 3b), and have detectable serum IFN- γ (not shown). These data reinforce

FIG. 4 Inflammatory response in the heart (a, c) and skeletal muscle (b, d) of $\beta_2m^{+/+}$ (a, b) and $\beta_2m^{-/-}$ (c, d) mice on day 18 of infection.

METHODS. Hearts and skeletal muscle were removed from mice and fixed in Bouin-Hollande's fixative for 48 h before routine processing and embedding in paraffin. Sections were stained with haematoxylin and eosin and evaluated by light microscopy on a single-blind basis as described previously²³. Arrows in c show the location of parasite-infected muscle cells (magnification, $\times 150$).



the finding that CD8⁺ T cells are not the major producers of IFN- γ in this infection⁵ and suggest that the effector function of CD8⁺ T cells in *T. cruzi* infection is unrelated to the capacity of this T-cell population to secrete IFN- γ . The enhanced IL-2 and IFN- γ production in the $\beta_2m^{-/-}$ mice may be a result of the increased proportion of CD4⁺ T cells in the spleen, but a similar increase in cytokine production was not observed in mice depleted of CD8⁺ T cells by antibody treatment^{4,5}.

A characteristic feature of *T. cruzi* infection is an acute and chronic phase tissue inflammation. In the acute phase, inflammation is seen in many different organs and may correlate with the presence of parasites. In the chronic stage of the infection, inflammation is restricted to muscle and is often seen in the presence of low levels of parasites¹¹⁻¹³. Heart and skeletal muscle tissue from $\beta_2m^{+/+}$ mice infected for 18 days display an inflammatory response typical of the acute stage of infection (Fig. 4a and b). However, $\beta_2m^{-/-}$ mice show a striking absence of inflammatory cells in both tissues (Fig. 4c and d), despite tissue parasite loads comparable to or higher than $\beta_2m^{+/+}$ mice. Hearts from $\beta_2m^{-/-}$ mice killed on day 30 of infection show extremely high parasite loads but only a moderate inflammatory response (not shown). Based on these results, death of the $\beta_2m^{-/-}$ mice may be the result of the direct effects of parasite destruction of host tissue or a toxemia resulting from the high parasite burden.

The $\beta_2m^{-/-}$ mutant mouse is an excellent system in which to evaluate the importance of antigen presentation to and the responsiveness of CD8⁺ T cells in resistance to infectious diseases. The lack of immune resistance to *T. cruzi* in β_2m -deficient mice provides additional documentation of the importance of CD8⁺ T cells in control of this intracellular protozoal infection. In addition to the CD8⁺ T cell deficiency in the $\beta_2m^{-/-}$ mice, these animals have also recently been shown to have impaired natural killer cell activity, a factor that could also increase their susceptibility to *T. cruzi* infection¹⁴. Interestingly, of the few infectious agents that have been evaluated in this mutant, *T. cruzi* is the only one causing the β_2m mutant to express this heightened degree of susceptibility. Experiments using influenza³, vaccinia (M. Steggs, personal communication) and mouse hepatitis viruses (B. H. Koller, unpublished) do not show that the β_2m mutant has a markedly enhanced susceptibility to these infectious agents. These results are contrary to expectations based on the demonstrated capacity of CD8⁺ T cells to destroy virus-infected cells and the relative paucity of data demonstrating an effector function for CD8⁺ T cells in *T. cruzi* infection. The protective immune response to *T. cruzi* clearly requires the combined efforts of a number of mechanisms, including CD4⁺ and CD8⁺ T cells, antibody production and natural killer activity^{4,15-18}. Disabling any one of these compartments could render mice unable to control the acute infection. Our results also suggest a relatively minor role for CD8⁺ T cells in immunoregulation and an important role for class I MHC gene products and/or CD8⁺ T cells in the acute inflammatory response in infected tissues. Previous studies have suggested that the acute-phase inflammation in *T. cruzi* infection is largely due to CD4⁺ T cells¹⁹⁻²². But we also find that infected tissues from mice in both the acute and chronic stages of infection contain abundant CD8⁺ T cells and that inflammation is moderated by the depletion of CD8⁺ T cells by antibody treatment (J. Sun and R.L.T., unpublished data). Further investigation of the relative contribution of lymphocyte subpopulations and their products to acute and chronic disease will be aided by the availability of other transgenic mouse strains containing gene disruptions. □

1. Tarleton, R. L. *J. Immunol.* **144**, 717-724 (1990).
2. Nabors, G. S. & Tarleton, R. L. *J. Immunol.* **146**, 3591-3598 (1991).
3. Tarleton, R. L. *J. Immunol.* **140**, 2769-2773 (1988).
4. Cunningham, D. S., Kuhn, R. E. & Rowland, E. C. *Infect. Immunol.* **23**, 155-160 (1978).
5. Ramos, C., Schadtler-Siwon, I. & Ortiz-Ortiz, L. *J. Immunol.* **122**, 1243-1247 (1979).
6. Scott, M. T. *Immunology* **44**, 409-417 (1981).
7. Beltz, L. A., Sonnenfeld, G. & Kierszenbaum, F. *Int. J. Parasit.* **19**, 555-559 (1989).
8. Koberle, F. *Adv. Parasit.* **6**, 63-116 (1968).
9. Andrade, Z. A. *Ciba Fdn Symp.* **99**, 214-233 (1983).
10. Santos-Buch, C. A. & Acosta, A. M. in *Immunology and Pathology of Trypanosomiasis* (ed. Tizard, I.) 145-183 (CRC, Boca Raton, 1985).
11. Liao, N.-S., Bix, M., Zijlstra, M., Jaenisch, R. & Raulet, D. *Science* **253**, 199-202 (1991).
12. Trischmann, T. M. *J. Immunol.* **130**, 1953-1957 (1983).
13. Krettl, A. U. *Memorias do Instituto Oswaldo Cruz* **79**(suppl.), 59-65 (1984).
14. Rottenberg, M., Cardoni, R. L., Andersson, R., Segura, E. L. & Orn, A. *Scand. J. Immunol.* **28**, 573-582 (1988).
15. Araujo, F. *Infect. Immunol.* **57**, 2246-2248 (1989).
16. Younes-Chennoufi, A. B., Said, G., Eisen, H., Durand, A. & Hontebeyrie-Joskowicz, M. *Trans. R. Soc. trop. Med. Hyg.* **82**, 84-89 (1988).
17. Ribeiro-dos-Santos, R., Pirmez, C. & Savino, W. *Res. Immunol.* **142**, 134-137 (1991).
18. Minoprio, P. *Res. Immunol.* **142**, 137-140 (1991).
19. Russo, M., Starobinas, N., Minoprio, P., Coutinho, A. & Hontebeyrie-Joskowicz, M. *Ann. Inst. Pasteur Immunol.* **139**, 225-236 (1988).
20. Postan, M., Cheever, A. W., Dvorak, J. A. & McDaniel, J. P. *Trans. R. Soc. trop. Med. Hyg.* **80**, 50-55 (1986).

ACKNOWLEDGEMENTS. We thank M. Heiges and S. Walker for technical assistance, J. Golden for help with the flow cytometric analysis, O. Smithies for research support and D. Krause and R. Damian for discussion of the manuscript. This work was supported by the NIH, the Cystic Fibrosis Foundation and the Georgia Affiliate of the American Heart Association.

C. elegans cell-signalling gene *sem-5* encodes a protein with SH2 and SH3 domains

Scott G. Clark*, Michael J. Stern*[†] & H. Robert Horvitz

Howard Hughes Medical Institute, Department of Biology, Room 56-629, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA

THE induction of the hermaphrodite vulva¹ and the migration of the sex myoblasts^{2,3} in the nematode *Caenorhabditis elegans* are both controlled by intercellular signalling. The gonadal anchor cell induces formation of the vulva from nearby hypodermal cells⁴, and a set of somatic gonadal cells attract the migrating sex myoblasts to their final positions². Many genes required for vulval induction have been identified¹, including the *let-23* receptor tyrosine kinase gene⁵⁻⁷ and the *let-60 ras* gene⁸⁻¹⁰. We report here the identification and characterization of a new gene, *sem-5* (*sem*, sex muscle abnormal), that acts both in vulval induction and in sex myoblast migration. On the basis of its DNA sequence, *sem-5* encodes a novel 228-amino-acid protein which consists almost entirely of one SH2 (SH, *src* homology region) and two SH3 domains. SH2 and SH3 domains are present in many signalling proteins regulated by receptor and non-receptor tyrosine kinases¹¹. Mutations that impair *sem-5* activity alter residues that are highly conserved among different SH2 and SH3 domains. Our results indicate that the *sem-5* gene encodes a novel protein that functions in at least two distinct cell-signalling processes.

We discovered *sem-5* in parallel mutant screens designed to identify genes involved either in vulval development or in sex myoblast migration. Three *sem-5* alleles (*n1619*, *n2019*, *n2030*) were isolated as suppressors of the multivulva (*Muv*) phenotype of the *lin-15(n765ts)* mutant (see legend to Table 1). Like mutations in the *let-23* and *let-60* genes isolated in this same screen (ref. 8; S.G.C. and H.R.H., manuscript in preparation), these *sem-5* mutations cause multiple recessive phenotypes: for example, many *sem-5* animals die as young larvae (Lethal phenotype), and many of those animals that survive exhibit a

Received 29 October 1991; accepted 21 January 1992.

1. Koller, B. H., Marrack, P., Kappler, J. W. & Smithies, O. *Science* **248**, 1227-1230 (1990).
2. Zijlstra, M. et al. *Nature* **344**, 742-746 (1990).
3. Eichelberger, M., Allan, W., Zijlstra, M., Jaenisch, R. & Doherty, P. C. *J. exp. Med.* **174**, 875-880 (1991).

* S.G.C. and M.J.S. contributed equally to this letter.

[†] Present address: Yale University, Boyer Center for Molecular Medicine, PO Box 9812, New Haven, Connecticut 06536-0812, USA.

vulvaless (Vul) phenotype (Table 1a). To establish the basis of the Vul phenotype, we determined the vulval cell lineages of a *sem-5* mutant (Table 1b). The three cells P5.p, P6.p and P7.p, which are normally induced by the anchor cell (ac) to express vulval cell lineages (1°, 2°), instead expressed a non-vulval cell lineage (3°) normally characteristic of the uninduced cells P3.p, P4.p and P8.p. The Vul phenotype of *sem-5* mutants is identical to that of animals lacking the inductive signal from the gonadal anchor cell⁴ (*ac*⁻ in Table 1b). Thus, like the *let-23* gene^{5,7} and the *let-60* gene^{8,9}, the *sem-5* gene is required for vulval induction as well as for an unknown process essential for early larval development.

Three other *sem-5* alleles (*n1779*, *n1781*, *n2195*) were isolated as suppressors of the sterile and 'clear' phenotype of the *clr-1*

1(*e1745ts*) mutant (see legend to Table 1). The *clr-1* gene (*clr*, clear, reflecting the clarity with which cell boundaries can be observed in these mutants¹²) seems likely to function in the regulation of sex myoblast migration because of its interactions with the gene *egl-15*, which acts in a cell signalling process that

TABLE 1 *sem-5* mutant phenotypes(a) Penetrances and *clr-1* suppressor activities

Allele	Lethal	Vul	Non-Vul	Mig	Non-Mig	<i>clr-1</i> suppression
Wild-type	0%	0%	100%	0%	100%	~
<i>n1619</i>	91%	9%	0%	19%	81%	++
<i>n2030</i>	71%	29%	0%	35%	65%	++
<i>n2019</i>	41%	58%	1%	17%	83%	++
<i>n1779</i>	0%	11%	89%	43%	57%	++
<i>n1781</i>	0%	1%	99%	0%	100%	+
<i>n2195</i>	0%	0%	100%	0%	100%	+

(b) P.n.p cell lineages

	P3.p	P4.p	P5.p	P6.p	P7.p	P8.p
Wild-type (<i>ac</i> ⁺)	3°	3°	2°	1°	2°	3°
Wild-type (<i>ac</i> ⁻)	3°	3°	3°	3°	3°	3°
<i>sem-5(n1619)</i>	3°	3°	3°	3°	3°	3°

a, Effects of the six *sem-5* alleles on viability, vulval development, sex myoblast (SM) migration and the *Clr-1* phenotype. Wild-type, the wild-type *C. elegans* Bristol strain N2 (ref. 32). Lethal, animals that arrested larval development and died as rod-like L1 or early L2 larvae. Vul, Vulvaless animals, which lacked a functional vulva. Non-Vul, Non-vulvaless animals, which had a functional vulva. For each strain, between 306 and 516 animals were scored. All Vul *n1619* and *n2030*, and more than 85% of Vul *n2019* hermaphrodites have a smooth ventral surface, which is characteristic of strong Vul mutants that seem to fail in or have only a very low level of vulval induction³³. Vul *n1779* and *n1781* hermaphrodites produce an abnormal vulva-like structure characteristic of weak Vul mutants that seem to be partially induced for vulval formation³³. Mig, Migration-defective sex myoblasts, which were displaced posteriorly by at least half the distance between the P6.p and P7.p cell nuclei³. Non-Mig, Sex myoblasts sex myoblast defective in migration. For each strain, between 16 and 46 individual sex myoblast positions were scored. Although the sex myoblast migration defect seen in *sem-5* mutants is less severe than that seen in *egl-15* mutants³, this defect is significant ($P < 0.05$) for all *sem-5* alleles except *n1781* and *n2195*. The statistical significances of differences between distributions of the sex myoblasts were determined using a one-tail unpaired *t*-test³. *clr-1* Suppression, Ability to suppress the sterile and 'clear' phenotype of *clr-1(e1745ts)* animals (see text). ++, Strong suppression, *clr-1* and *sem-5* double mutant animals are fertile and healthy; +, weak suppression, *clr-1* and *sem-5* double mutant animals are fertile yet display a weak clear phenotype; -, no suppression. All phenotypes were scored at 20 °C except for *clr-1(e1745ts)* suppression, which was scored at 25 °C. We isolated the mutations *n1619*, *n2019* and *n2030* as suppressors of the multivulva (Muv) phenotype of the *lin-15(n765ts)* mutant after mutagenesis with ethyl methanesulphonate³² and the examination of the progeny of 38,000 F₁ animals⁸. We isolated the mutations *n1779*, *n1781* and *n2195* as suppressors of the sterile and clear phenotype of the *clr-1(e1745ts)* mutant after mutagenesis with ethyl methanesulphonate and the examination of the progeny of 118,000 F₁ animals. This latter screen was designed to identify mutations that, like some mutations in the gene *egl-15*, both interfered with sex myoblast migration and suppressed the effects of the *clr-1(e1745ts)* mutation (M.J.S. and H.R.H., manuscript in preparation). From mapping and complementation experiments, we found that all six of these new mutations are allelic. Scoring both the Vul and Lethal phenotypes, we mapped the *n1619* and *n2030* mutations between the *unc-10* and *xol-1* genes on the X chromosome (see Fig. 1a). For *n1619* and *n2030*, respectively, of the Unc non-Dpy progeny of *sem-5/unc-10(e102) xol-1(y9) dpy-6(e14)* hermaphrodites, 14/20 and 17/30 segregated Vul, Lethal and non-Xol progeny, 1/20 and 2/30 segregated non-Vul, non-Lethal and non-Xol progeny, and 5/20 and 11/30 segregated Xol, non-Vul and non-Lethal progeny. Scoring the *clr-1* suppression phenotype, we mapped the *n1779* mutation to the same interval. For *n1779*, of the Unc non-Dpy progeny of *clr-1(e1745ts)/clr-1(e1745ts); sem-5/unc-10(e102) xol-1(y9) dpy-6(e14)* hermaphrodites, 46/73 segregated non-Clr (suppressed for clear phenotype) non-Xol progeny, 3/73 segregated non-Clr and non-Xol progeny, and 24/73 segregated Xol Clr progeny. *n1619*, *n1781*, *n2019*, *n2030* and *n2195* all fail to complement *n1779* for the *clr-1* suppression phenotype, and *n1619*, *n2019* and *n2030* all fail to complement *n1779* for the vulvaless and sex myoblast migration-defective phenotype. b, Vulval cell lineages of wild-type (strain N2) (*ac*⁺), wild-type without the anchor cell (*ac*⁻) and *sem-5(n1619)* Vul animals. In normal (*ac*⁺) wild-type animals, the six cells P3.p-P8.p constitute the vulval equivalence group, and each cell expresses one of three distinct cell lineages, known as 1°, 2° and 3°; the descendants of the 1° and 2° lineages generate the vulva, whereas the descendants of the 3° lineage are non-vulval (reviewed in ref. 2). In wild-type animals lacking the gonadal anchor cell (*ac*⁻), all six cells express the 3° lineage and no vulva is formed⁴. In the eight *sem-5(n1619)* animals we examined, almost all of the P3.p-P8.p cells expressed the 3° lineage, indicating that in these animals the pathway of vulval induction was blocked. The exceptions were two P6.p cells that expressed an apparent 1° lineage, the normal fate of P6.p; such incomplete penetrance is typical of many Vul mutants⁵. Cell lineages were determined by direct observation using Nomarski optics, as described previously³⁴.

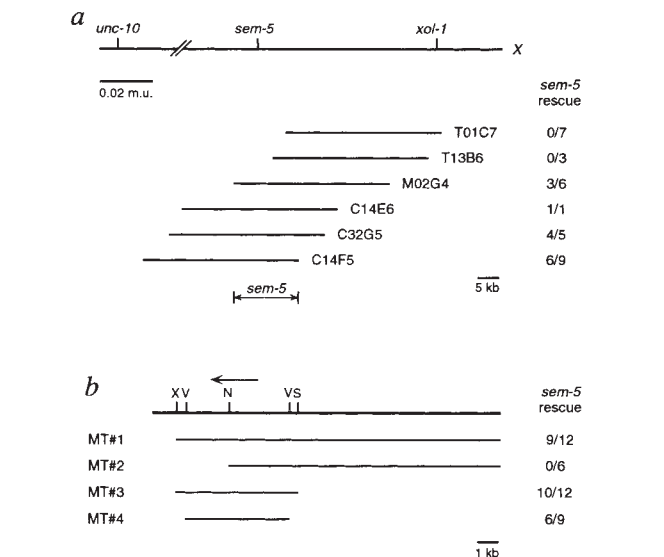


FIG. 1 Genetic and physical maps of the *sem-5* region of the X chromosome. a, A genetic map of the *sem-5* region with cosmid clones representing this region depicted below (see Table 1 legend for map data). Cosmid clones were tested for the rescue of the vulvaless (Vul) phenotype of *sem-5(n2019)* animals when maintained as a transmitted extrachromosomal array after germline transformation¹⁵. The fraction of independently derived transformed F₂ populations rescued for the Vul phenotype of *sem-5(n2019)* animals is presented for each cosmid clone. These results indicate that the *sem-5*-rescuing activity is contained in the region marked. b, A restriction map of the right end of the *sem-5*-rescuing cosmid C14F5 and structures of genomic subclones. Subclones were similarly assayed for rescue of the Vul phenotype of *sem-5(n2019)* animals. The 6-kb genomic fragment MT#3 was also shown to rescue the *clr-1* suppression phenotype of *sem-5(n1779)* animals. The 5-kb genomic fragment MT#4 also rescues the lethal phenotype of *sem-5(n2019)* animals and the Vul, Lethal and *clr-1* suppression phenotypes of *sem-5(n2030)* animals. Thus, *sem-5*-rescuing activity is contained in the 5-kb *EcoRV* fragment, MT#4. The *sem-5* transcript and direction of transcription are represented by the arrow (see data in Fig. 2). Restriction site abbreviations are V (*EcoRV*), N (*NheI*), S (*SacI*), and X (*XhoI*).

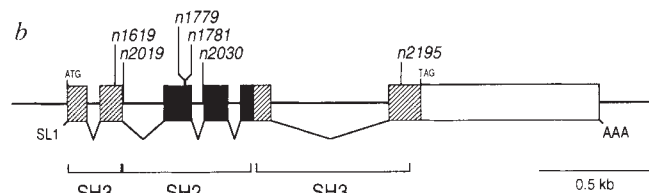
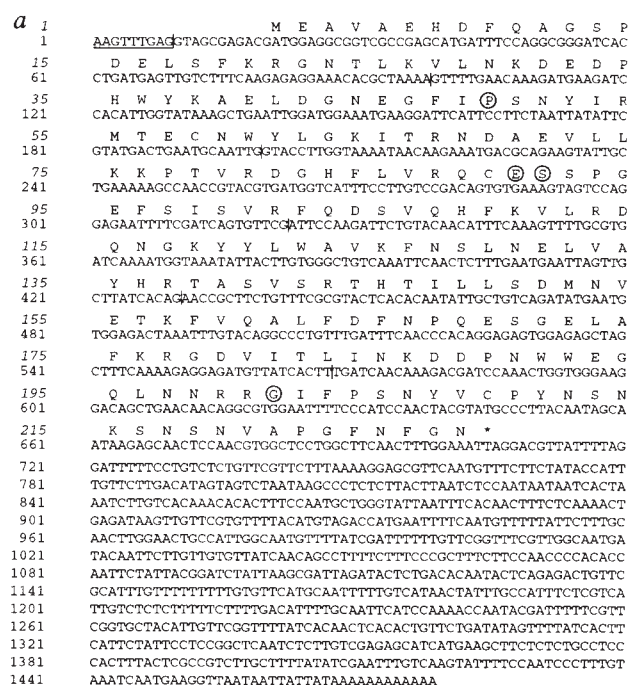
METHODS. Germline transformation¹⁵ of *sem-5(n2019)* mutants was done by co-injecting test DNA (10 µg ml⁻¹ and the dominant *rol-6(su1006)* roller marker (plasmid pRF4 at 80 µg ml⁻¹). Transgenic animals typically carry co-injected DNAs as an extrachromosomal array¹⁵ and are identified by the roller phenotype (Rol) conferred by pRF4. Roughly 99% of surviving *sem-5(n2019)* animals are vulvaless (Table 1a and data not shown). A *sem-5*-rescued F₂ population was defined as an F₂ roller population (derived from F₁ roller animals) with at least 50% of the animals non-Vul. Genomic subclones that rescued the Vul phenotype of *sem-5(n2019)* animals in F₂ populations also rescued the Vul phenotype of the F₁ Rol progeny of injected animals (for MT#1, 18% of F₁ roller animals were non-Vul; MT#3, 38%; MT#4, 70%). To test rescue of the lethal phenotype, we counted the progeny of Rol non-Vul transgenic animals to determine the numbers of viable and of rod-like larval lethal progeny. To test rescue of the *clr-1* suppression phenotype, we co-injected *clr-1(e1745ts)/sem-5* animals with *sem-5* genomic DNA and pRF4 as described above, and grew progeny at 15 °C. Transgenic Rol animals were shifted to 25 °C and observed for the Clr phenotype, which is normally suppressed by mutations in *sem-5*.

controls sex myoblast migration (ref. 3; M.J.S. and H.R.H., manuscript in preparation). Like *egl-15* mutants, *sem-5* mutants have posteriorly displaced sex myoblasts (Table 1a) and both *egl-15* and *sem-5* mutations can affect the communication between the sex myoblasts and the gonad (ref. 3; M.J.S. and H.R.H., manuscript in preparation). All six *sem-5* alleles suppress the sterile phenotype of the *clr-1(e1745ts)* mutant; all but *n1781* and *n2195* cause a defect in sex myoblast migration. Only the three alleles (*n1619*, *n2019*, *n2030*) isolated as *lin-15(n765ts)* suppressors cause a larval lethal and highly penetrant vulvaless phenotype. These results indicate that the *sem-5* gene is required for the regulation of sex myoblast migration as well as for vulval induction and early larval development.

We cloned the *sem-5* gene based upon its map position between the genes *unc-10* and *xol-1* on the X chromosome, roughly 0.07 map units left of *xol-1* (Fig. 1a; see also Table 1 legend). Using the nearly complete physical map of the *C. elegans* genome, which consists of overlapping sets of cosmids and yeast artificial chromosomes (refs 13, 14; A. Coulson and J. Sulston, personal communication), we obtained cosmid clones representing the region left of a *xol-1* mutation (L. Miller and

B. Meyer, personal communication). We tested these cosmids in germ-line transformation experiments¹⁵ and identified four overlapping cosmids (C14F5, C32G5, C14E6, M02G4) that rescued the Vul phenotype of *sem-5(n2019)* animals (Fig. 1a). A 5-kilobase (kb) genomic fragment (MT#4) derived from the region common to these four cosmids rescued both the Vul phenotype of *sem-5(n2019)* animals and also the Vul, Lethal and *clr-1* suppression phenotypes of *sem-5(n2030)* animals (Fig. 1b). This 5-kb fragment detected a single 1.5-kb transcript in mixed stage poly(A)⁺ RNA from wild-type (strain N2) animals (data not shown). These experiments suggest that the *sem-5* gene is in the 5-kb genomic fragment and might encode a 1.5-kb transcript.

We determined the DNA sequences of this 5-kb genomic fragment and of two independently derived complementary DNAs from this region (Fig. 2). These cDNAs were 1,479 and 1,447 nucleotides long, consistent with the size of the 1.5-kb presumptive *sem-5* messenger RNA. Both cDNAs contained nine nucleotides of the 22-nucleotide SL1 *trans*-spliced leader sequence¹⁶ and the longer one contained a 12-nucleotide poly(A) tail, indicating that this cDNA represents a full-length *sem-5*



C	Allele	Wild-type sequence	Mutant sequence	Substitution
	n1619	CCT	CTT	P → L
	n1779	GAA	ΔAA	E → K
	n1781	AGT	ΔAT	S → N
	n2019	TTG gtatgttc	TTG atgtgttc	
	n2030	ttttccag ATT	ttttccaa ATT	
	n2195	GGA	ΔGA	G → R

FIG. 2 a, Nucleotide and predicted amino-acid sequences (single-letter code) corresponding to the *sem-5* cDNAs. The nine nucleotides derived from the 22-nucleotide SL1 *trans*-spliced leader¹⁶ are underlined. Nucleotides are numbered on the left beginning at the first nucleotide present in the cDNAs, which is within the SL1 *trans*-spliced leader sequence. Amino acids are numbered in italics on the left beginning at the first predicted Met. The positions of splice sites are marked by vertical lines in the nucleotide sequence. The amber termination codon is denoted by an asterisk. Residues altered in ethyl methanesulphonate-induced mutants are circled: *n1619* (P49L, substitution of L for P at codon 49), *n1779* (E90K), *n1781* (S91N) and *n2195* (G101R.) b, *sem-5* genomic structure derived from cDNA and genomic sequences. The SH2 region is shown as solid boxes, the SH3 regions are shown as cross-hatched boxes and the 3' untranslated region is shown as an open box. The boundaries of the SH2 and SH3 domains are marked by brackets. 'SL1' denotes the *trans*-spliced SL1 leader sequences; 'ATG' and 'TAG' show the predicted start and stop sites of translation, respectively. The locations of the six *sem-5* mutations are shown. c, *sem-5* mutants. All six *sem-5* mutations are G:C to A:T transitions. *n1619*, *n1779*, *n1781* and *n2195* are missense mutations. The codon altered and the expected amino-acid substitution for each mutant is shown. *n2019* and *n2030* alter invariant nucleotides in splice donor and acceptor consensus sequences, respectively. The wild-type and mutant sequences are shown with exon

sequences in upper case and intron sequences in lower case letters. The *C. elegans* 5' and 3' consensus splice sites are RAG|gtaagttt and wwtttcag|NNN, respectively²⁴ (R is A or G; W is A or T; N is A, C, G or T). METHODS. The 6-kb genomic fragment from MT#3 was used to screen about 400,000 plaques of a λ ZAP cDNA library derived from mixed stage poly(A)⁺ RNA³⁵. Ten positive clones were identified, of which eight had a 1.5-kb insert, one had a 1.4-kb insert and one had a 1.3-kb insert. The shorter clones seemed, by restriction map analysis, to be non-full-length *sem-5* cDNAs. The 5-kb genomic *EcoRV* *sem-5* fragment was subcloned into the *EcoRV* site of pBluescript KS(+) (Stratagene) in both orientations. We determined the complete DNA sequences of both strands of the genomic fragment and of the opposite strands of two different 1.5-kb cDNAs, MT#5 and MT#6, using Sequenase 2.0 (USB), essentially according to instructions provided by the manufacturer. The sequences of both cDNAs were identical, except that MT#6 ended 20 nucleotides before the 12 nucleotide poly-(A) tail present in MT#5. Exons of mutant genomic DNA were amplified by asymmetric polymerase chain reaction (PCR)³⁶ using primers that flanked the open reading frames (ORF). The primers used were as follows: exons 1–5, 5'-AGATACAGTCTCTCAAGCAGCG and 5'-CGAAATACAGTCCAAGATAT-ATCC; exon 6, 5'-CGGTGATCATCTATTGTGCGG and 5'-TGGAGATTAAGTAAG-AGAGGGC. Preparation of genomic DNA from five gravid adult hermaphrodites and subsequent amplification by PCR were essentially as described⁸. The DNA sequences were determined using primers internal to the PCR products. The sequencing primers used were as follows: exons 1–5, 5'-CAATGTACTCTGTTATGT, 5'-GATCTTACTTCTCGCGG 5'-CACAGGTGAG-TACAC, 5'-TCAAACCTTCAATTGAAC, 5'-GTTGAATTTGACAGCG and 5'-CGCGGAGAAGTAAGATCGACTG; exon 6, 5'-GTTGAGCAACCCCGAA and 5'-GAAACATTGAACGCTCC. We determined the DNA sequences of both strands of the ORFs of all six exons and of at least 5 base pairs beyond the 5' and 3' ends of each ORF for all six mutations. The search of GenBank, PIR and SWISS-PROT databases was done at the National Center for Biotechnology Information (NCBI) using the BLAST network service.

protein, Sem-5 presumably interacts with phosphotyrosine-containing proteins, perhaps Let-23 or its substrates, to effect signal transduction. Our results suggest that Sem-5 and similar molecules are required for signalling between protein tyrosine kinases and Ras proteins. In addition, because the *sem-5* gene acts in both vulval induction and sex myoblast migration, it seems likely that a tyrosine kinase-mediated pathway also con-

trols sex myoblast migration. Further study of these two processes and analysis of the interactions of *sem-5* with both common and distinct components of these two pathways should provide insight into the basic mechanisms of receptor tyrosine kinase action. We hope that such studies will reveal how these types of signal transduction pathways control cell fates and cell movements during animal development. □

Received 14 February; accepted 6 March 1992.

- Horvitz, H. R. & Sternberg, P. W. *Nature* **351**, 535–541 (1991).
- Thomas, J. H., Stern, M. J. & Horvitz, H. R. *Cell* **62**, 1041–1052 (1990).
- Stern, M. J. & Horvitz, H. R. *Development* **113**, 797–803 (1991).
- Kimble, J. *Dev Biol.* **87**, 286–300 (1981).
- Ferguson, E. L., Sternberg, P. W. & Horvitz, H. R. *Nature* **326**, 259–267 (1987).
- Aroian, R. V., Koga, M., Mendel, J. E., Ohshima, Y. & Sternberg, P. W. *Nature* **348**, 693–699 (1990).
- Aroian, R. V. & Sternberg, P. W. *Genetics* **128**, 251–267 (1991).
- Beitel, G. J., Clark, S. G. & Horvitz, H. R. *Nature* **348**, 503–509 (1990).
- Han, M., Aroian, R. V. & Sternberg, P. W. *Genetics* **126**, 899–913 (1990).
- Han, M. & Sternberg, P. W. *Cell* **63**, 921–931 (1990).
- Koch, C. A., Anderson, D., Moran, M. F., Ellis, C. & Pawson, T. *Science* **252**, 668–674 (1991).
- Hedgecock, E. M., Culotti, J. G. & Hall, D. H. *Neuron* **4**, 61–85 (1990).
- Coulson, A., Sulston, J., Brenner, S. & Karn, J. *Proc. natn. Acad. Sci. U.S.A.* **83**, 7821–7825 (1986).
- Coulson, A., Waterston, R., Kiff, J., Sulston, J. & Kohara, Y. *Nature* **335**, 184–186 (1988).
- Mello, C. C., Kramer, J. M., Stinchcomb, D. & Ambros, V. *EMBO J.* **10**, 3959–3970 (1991).
- Krause, M. & Hirsh, D. *Cell* **49**, 753–761 (1987).
- Altshul, S. F. & Lipman, D. J. *Proc. natn. Acad. Sci. U.S.A.* **87**, 5509–5513 (1990).
- Sadowski, I., Stone, J. C. & Pawson, T. *Molec. cell Biol.* **6**, 4396–4408 (1986).
- Stahl, M. L., Ferenz, C. R., Kelleher, K. L., Kriz, R. W. & Knopf, J. L. *Nature* **332**, 269–272 (1988).
- Suh, P. G., Ryu, S. H., Moon, K. H., Suh, H. W. & Rhee, S. G. *Cell* **54**, 161–169 (1988).
- Vogel, U. S. *et al.* *Nature* **335**, 90–93 (1988).
- Mayer, B. J., Hamaguchi, M. & Hanafusa, H. *Nature* **332**, 272–275 (1988).
- Lehmann, J. M., Riethmüller, G. & Johnson, J. P. *Nucleic Acids Res.* **18**, 1048 (1990).
- Fields, C. *Nucleic Acids Res.* **18**, 1509–1512 (1990).

- Mount, S. M. *Nucleic Acids Res.* **10**, 459–472 (1982).
- Shivelman, E., Lifshitz, B., Gale, R. P., Roe, B. A. & Canaani, E. *Cell* **47**, 277–284 (1986).
- Mayer, B., Jackson, P., Van Etten, R. & Baltimore, D. *Molec. cell Biol.* **12**, 609–618 (1992).
- Moran, M. F. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **87**, 8622–8626 (1990).
- Matsuda, M., Mayer, B. J. & Hanafusa, H. *Molec. cell Biol.* **11**, 1607–1613 (1991).
- Rodaway, A. R., Sternberg, M. J. & Bentley, D. L. *Nature* **342**, 624 (1989).
- Matsuda, M., Mayer, B. J., Fukui, Y. & Hanafusa, H. *Science* **248**, 1537–1539 (1990).
- Brenner, S. *Genetics* **77**, 71–94 (1974).
- Ferguson, E. L. & Horvitz, H. R. *Genetics* **110**, 17–72 (1985).
- Sulston, J. E. & Horvitz, H. R. *Dev Biol.* **56**, 110–156 (1977).
- Barstead, R. J. & Waterston, R. H. *J. Biol. Chem.* **264**, 10177–10185 (1989).
- McCauley, P. C. in *PCR Protocols* (eds Innis, M. A., Gelfand, D. H., Sninsky, J. J. & White, T. J.) 76–83 (Academic, San Diego, 1990).
- Jung, G., Korn, E. D. & Hammer, J. A. III *Proc. natn. Acad. Sci. U.S.A.* **84**, 6720–6724 (1987).
- Katzav, S., Martin, Z. D. & Barbacid, M. *EMBO J.* **8**, 2283–2290 (1989).
- Drubin, D. G., Mulholland, J., Zhu, Z. M. & Botstein, D. *Nature* **343**, 288–290 (1990).
- Takeya, T. & Hanafusa, H. *Cell* **32**, 881–890 (1983).
- Wasenius, V. M. *et al.* *J. Cell Biol.* **108**, 79–93 (1989).
- Huang, C. C., Hammond, C. & Bishop, J. M. *J. molec Biol.* **181**, 175–186 (1985).

ACKNOWLEDGMENTS. We thank N. Sotocodenhia for isolating *sem-5*(n2195), L. Miller and B. Meyer for physical map data for *xol-1*(-49), R. Barstead for the cDNA library, C. Mello for the plasmid pRF4, J. Sulston and A. Coulson for the physical map of the *sem-5* region, and C. Bargmann, J. Kaplan and E. Jorgensen for their comments. This work was supported by a research grant from the US Public Health Service to H.R.H. and by a postdoctoral fellowship from the American Cancer Society to M.J.S., H.R.H. is an Investigator and M.J.S. was a Research Associate of the Howard Hughes Medical Institute.

Action of brefeldin A blocked by activation of a pertussis-toxin-sensitive G protein

Nicholas T. Ktistakis*, Maurine E. Linder† & Michael G. Roth*

Departments of *Biochemistry and †Pharmacology, University of Texas Southwestern Medical Center at Dallas, 5323 Harry Hines Boulevard, Dallas, Texas 75235-9038, USA

In many mammalian cells brefeldin A interferes with mechanisms that keep the Golgi apparatus separate from the endoplasmic reticulum^{1–8}. The earliest effect of brefeldin A is release of the coat protein β -COP from the Golgi^{9–11}. This release is blocked by pretreatment with GTP- γ S or AIF₄⁻ (ref. 12). The AIF₄⁻ ion activates heterotrimeric G proteins¹³ but not proteins of the *ras* superfamily¹⁴, suggesting that a heterotrimeric G protein might control membrane transfer from the endoplasmic reticulum to the Golgi. We report here that mastoparan, a peptide that activates heterotrimeric G proteins^{15,16}, promotes binding of β -COP to Golgi membranes *in vitro* and antagonizes the effect of brefeldin A on β -COP in perforated cells and on isolated Golgi membranes. This inhibition is greatly diminished if cells are pretreated with pertussis toxin before perforation. Thus, a heterotrimeric G protein of the G_i/G_o subfamily regulates association of coat components with Golgi membranes.

To test the effects of activators of trimeric G proteins on the action of brefeldin A (BFA), we used a perforated cell system¹⁷ that allows cytosolic proteins to be completely washed away from cellular membranes¹⁸. In this reconstituted system, exogenously added cytosol, GTP and an ATP-generating system are necessary to maintain β -COP on membranes or to bind β -COP back to membranes of perforated cells from which it has been removed by energy depletion and low-concentration Mg²⁺ washes. Pretreatment with GTP- γ S antagonizes the release of β -COP by 36 μ M BFA only if cells are treated with ≥ 40 μ M GTP- γ S (Fig. 1a, e). Pretreatment with BFA releases β -COP

from the Golgi and prevents β -COP in the cytosol from reassembling on Golgi membranes to which GTP- γ S has been added. But small, punctate, perinuclear accumulations of β -COP form when cells pretreated with 36 μ M BFA are later treated with cytosol and 150 μ M GTP- γ S (Fig. 1f).

To determine whether the GTP- γ S effect on BFA requires a component in the cytoplasm, on membranes, or both, cells were depleted of ATP before perforation and washed extensively in low-concentration Mg²⁺ buffer after perforation to remove β -COP from the cells. Cells depleted of β -COP were treated for 5 min with GTP- γ S, and then washed extensively (Fig. 2a). Cells were then incubated in cytosol for 10 min to allow β -COP to bind and then 18 μ M BFA was added for an additional 10 min to half the samples. In both BFA-treated and untreated cells, β -COP formed a perinuclear concentration typical of Golgi binding in the GTP- γ S-treated perforated cells (Fig. 2c, e). The complementary experiment is shown in Fig. 2b, d, f. Cytosol treated with 1 mM GTP- γ S, and then dialysed, allowed binding of β -COP on the Golgi of perforated cells in the absence of BFA (Fig. 2d), but failed to prevent the release of β -COP from Golgi membranes of cells later treated with the drug (Fig. 2f). Thus, GTP- γ S prevents BFA from releasing β -COP by acting on a membrane-associated, and not a cytosolic component of the perforated cells.

The aluminium tetrafluoride (AIF₄⁻) ion, which specifically activates heterotrimeric G proteins, prevents the action of BFA on the Golgi¹², suggesting that the Golgi membrane component sensitive to GTP- γ S in our experiments might be such a protein. Thus, perforated cells or isolated Golgi membranes were treated with active mastoparans (Mas1 and Mas7) that mimic the ability of receptors to catalyse nucleotide exchange on G protein α subunits (G_s) (refs 15, 16). As a control, two weak analogues of mastoparan, Mas11 and Mas17 (ref. 16), were also used. As shown in Fig. 3a, the mastoparan analogue Mas7 (ref. 16) prevented BFA from releasing β -COP from perforated cells. Mas11 and Mas17 had minimal effect, indicating that the detergent properties of mastoparan were not important for preventing release of β -COP from the Golgi (data not shown). The degree to which mastoparan analogues stimulated the binding of β -COP to Golgi membranes was measured by western blotting

using the assay of Orci *et al.*¹⁹ (Fig. 3b). Golgi in samples treated with the active mastoparans (Fig. 3b, lanes 5 and 7) bound 10-fold more β -COP than the Golgi in samples treated with cytosol alone (lane 1), or with the inactive Mas17 (not shown), and fivefold more than in samples treated with GTP- γ S (lane 3). BFA inhibited 60% of the binding of cytosolic β -COP to Golgi membranes (lane 2). BFA action was inhibited by pretreatment with GTP- γ S, Mas1, or Mas7 (lanes 3–8), but not with Mas17 (not shown). Golgi membranes treated with Mas1 or with Mas7 bound 48% or 52% of all the β -COP present in the cytosol used in the assay. In the absence of added membranes, 7% of β -COP from cytosol treated with Mas7 was found in the pellet fraction. Binding of β -COP to Golgi membranes in the presence of Mas1 or Mas7 increased with increasing amounts of cytosol in the concentration range used in our experiments, but was inhibited at high ratios of cytosol to Golgi membranes.

ADP-ribosylation of G_{α} by pertussis toxin prevents mastoparan-stimulated guanine nucleotide exchange measured *in vitro*¹⁵. Thus cells were treated with pertussis toxin overnight before perforation, and then assayed for the ability of GTP- γ S or mastoparan to block the release of β -COP from the Golgi by BFA. Tight, perinuclear, very brightly staining concentrations of β -COP were observed when perforated cells were incubated in the reconstitution buffer after treatment with GTP- γ S, regardless of whether the cells had been treated with pertussis toxin overnight (Fig. 4a, b). By contrast, perforated cells treated with pertussis toxin, mastoparan and then BFA showed greatly reduced perinuclear accumulation of β -COP (compare Fig. 4c

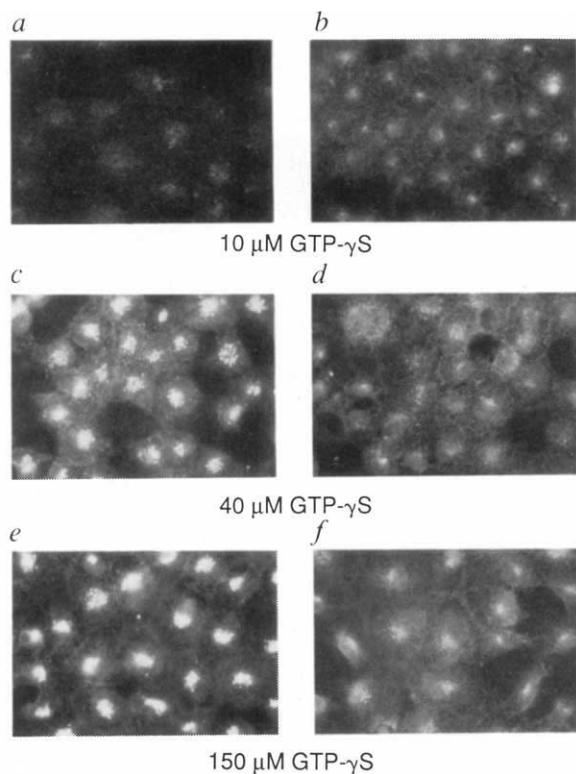


FIG. 1 In all experiments cells were perforated and cytosol containing an ATP-regenerating system was prepared as described¹⁸. After the perforated cells were washed, cytosol and GTP- γ S were added back to half the samples for 10 min, followed by the addition of 36 μ M BFA for 10 min (a, c, e). Cytosol containing 36 μ M BFA was added to the other half of the samples for 10 min, followed by the addition of GTP- γ S (b, d, f). Cells were then fixed and stained for immunofluorescence with an anti- β -COP monoclonal antibody as described¹⁸. The bright concentration of perinuclear immunofluorescence at the centre of the cell corresponds to the position of the Golgi complex in these cells¹⁸. β -COP remained on Golgi membranes in the presence of BFA if sufficient GTP- γ S was added before BFA (e).

and d), indicating that pertussis toxin diminished the ability of mastoparan to inhibit BFA.

Pretreatment of Golgi membranes and cytosol *in vitro* with $\beta\gamma$ subunits of trimeric G proteins prevents GTP- γ S-enhanced binding of β -COP to Golgi membranes to an extent comparable to BFA pretreatment²⁰. Our results indicate that activation of a pertussis-toxin-sensitive G_{α} promotes the binding of β -COP to Golgi membranes and antagonizes the action of BFA. We have identified a pertussis-toxin-sensitive G_{α} protein(s) on isolated Golgi cisternae both by western blotting and by immunocytochemistry (unpublished results). The major G_{α} detected has an apparent relative molecular mass of 41,000 (M_r 41K) and its reactivity with antisera suggests that it may be $G_{i\alpha 1}$. Other pertussis-toxin-sensitive G_{α} s may be present in lower abundance, including $G_{i\alpha 3}$, which was recently identified on the Golgi complex of LLC-PK₁ cells and has an inhibitory effect on secretion²¹. We note that the PtK1 cell line which is resistant to BFA owing to a nondiffusible factor located on the Golgi complex¹⁸ contains a G_{α} with an apparent M_r of 39K which is absent from all of the BFA-sensitive cells that we have investigated (unpublished results). Thus, the resistance of PtK1

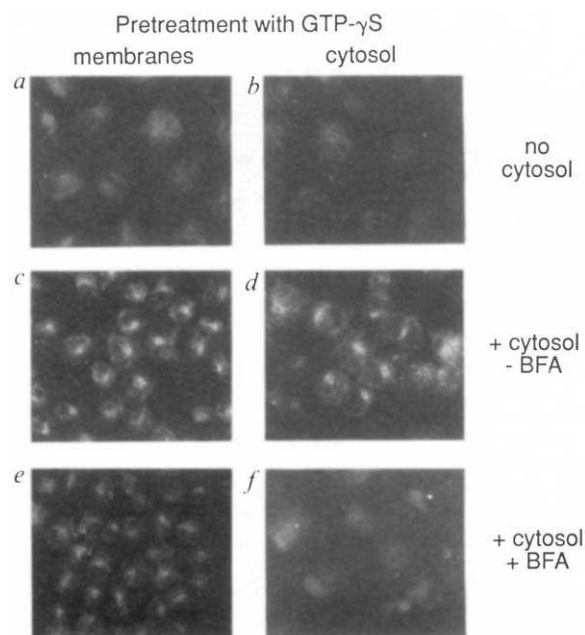


FIG. 2 Pretreatment of membranes, but not cytosol, with GTP- γ S prevents the action of BFA. MA104 cells grown on coverslips were incubated in medium containing 50 mM 2-deoxyglucose and 0.02% Na₂S₂O₈ in PBS for 5 min to deplete cellular ATP levels, resulting in the release of β -COP from Golgi membranes. Cells were then perforated with nitrocellulose¹⁷ and incubated in 90 mM KCl + 50 mM HEPES, pH 7.2, for 5 min, then washed for 20 s in 100 ml of the same buffer. Perforated cells depleted of β -COP were treated with 1 mM GTP- γ S for 5 min at 37 °C (a, c, e). Cells were washed extensively to remove GTP- γ S, and a sample was fixed and stained (a). Cytosol was added to the cells for 10 min and then BFA at 18 μ M was added for a second 10 min before cells were fixed and analysed by immunofluorescence. β -COP from the cytosol bound to the Golgi of cells treated with GTP- γ S and was not released by BFA (compare c and e). Cytosol (1 ml at a protein concentration of 5 mg ml⁻¹ containing 10 μ M GTP- γ S-binding sites) was treated with 1 mM GTP- γ S at 37 °C for 1 h and then dialysed overnight at 4 °C. After perforated cells were depleted of β -COP (b), GTP- γ S-treated cytosol was added for 10 min followed by the addition of 18 μ M BFA to half the samples for an additional 10 min and cells were fixed and stained with anti- β -COP monoclonal antibody. In the absence of BFA, β -COP from cytosol pretreated with GTP- γ S bound to Golgi membranes, as indicated by the concentration of perinuclear fluorescence in every cell (d). In the presence of BFA (f), weaker fluorescence was observed throughout each cell, which resembled the starting membrane preparation with little or no perinuclear concentration.

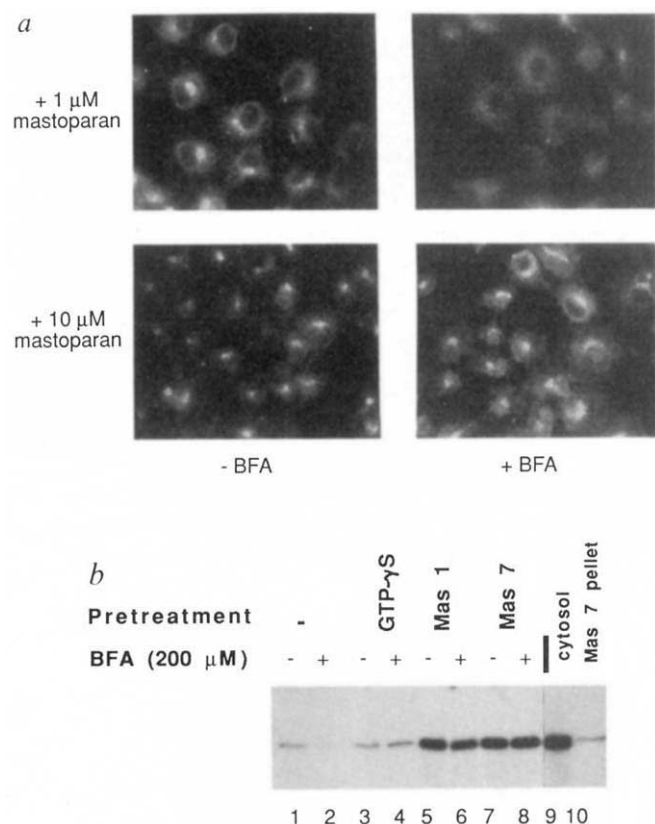


FIG. 3 A mastoparan analogue (Mas7) plus GTP promotes the binding of β -COP to Golgi membranes and antagonizes its release by BFA. **a**, Perforated cells depleted of Golgi coat proteins were prepared as for Fig. 2. Cytosol containing 1 mM GTP, an ATP-regenerating system, and Mas7 at the concentrations shown was added for 10 min. BFA was added at a final concentration of 3.6 μM for an additional 10 min. β -COP from the cytosol bound to the Golgi of cells in the presence of 10 μM Mas7 and was not released by BFA (compare bottom two panels). **b**, Effect of mastoparans on binding β -COP to Golgi membranes *in vitro*. Golgi membranes from MA104 cells were incubated for 5 min with MA104 cytosol, an ATP-regenerating system, GTP, and the indicated additives (lanes 1–8). At the end of 5 min, samples received 200 μM BFA as indicated, and all samples were incubated at 37 $^{\circ}\text{C}$ for a further 5 min. Golgi membranes were pelleted for 6 min at 14,000g, resolved by SDS-PAGE, transferred to nitrocellulose and probed with anti- β -COP antibodies. Lane 9 shows the amount of β -COP in the volume of cytosol used for every incubation in lanes 1–8. Lane 10 shows the amount of β -COP that pelleted from cytosol incubated with MAS7, no BFA and with Golgi membranes omitted. In experiments like the one shown, the percentage of β -COP remaining on Golgi membranes after BFA treatment was $43 \pm 8\%$ with no pretreatment, $80 \pm 11\%$ with GTP- γS , $93 \pm 3\%$ with Mas7, 78% with Mas1, and $40 \pm 5\%$ with Mas17. Standard deviations from four separate experiments are given.

METHODS. The assay is adapted from ref. 19. Golgi membranes were isolated from MA104 cells using sedimentation of post-nuclear supernatants on discontinuous sucrose gradients as described by Balch *et al.*²². Membranes at the 0.8–1.2M sucrose interphase were collected in a minimum volume and stored at -70°C . By electron microscopy, this fraction contained the majority of intact Golgi stacks. Cytosol was isolated from MA104 cells and is defined as the supernatant of 150,000g spin. This cytosol was dialysed overnight against two changes of 1,000 volumes each of 50 mM HEPES, 90 mM KCl, pH 7.2, and was stored at -70°C . For the incubations shown above, the following components were present in a final volume of 400 μl : 50 mM HEPES, pH 7.2, 78 mM KCl, 2.5 mM MgCl_2 , 1 mM ATP, 1 mM GTP, 2 mM creatine phosphate, 7 U creatine phosphokinase, 20 μl cytosol (at 1–2 mg ml^{-1}), and 10 μl Golgi membranes (at 0.3–0.5 mg ml^{-1}). Reagents added to the standard assay are indicated and were present in the following concentration: GTP- γS , 100 μM ; Mas1 or Mas7, 25 μM ; BFA, 200 μM . All samples contained 1% dimethyl sulphoxide. For western blots, the primary antibody against β -COP, ascites fluid M3A5 (ref. 23), was used at 1:500 dilution, and was followed by anti-mouse IgG and ^{125}I -labelled anti-rabbit Fab fragments.

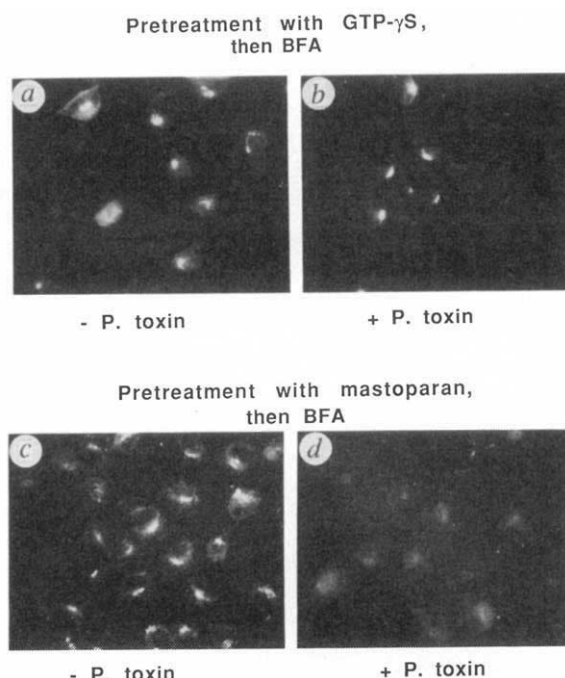


FIG. 4 Pertussis toxin diminishes the Mas7 inhibition of BFA. Living cells were treated overnight with 100 ng ml^{-1} pertussis toxin in cell culture medium (resulting in some cell rounding). Perforated cells depleted of Golgi coat proteins were prepared from toxin-treated or -untreated cells as before. Half of the cells were treated with GTP- γS as in Fig. 2 and half were treated with 10 μM Mas7, as in Fig. 3, before challenge with 3.6 μM BFA. BFA failed to release β -COP from the Golgi of cells pretreated with GTP- γS , regardless of whether they had been treated with pertussis toxin (**a**, **b**). But the ability of Mas7 to inhibit the action of BFA was greatly diminished, although not totally lost, when cells were previously treated with pertussis toxin (**c**, **d**). Pertussis toxin treatment of the cells *in vivo* was successful as a membrane preparation from cells treated in the same experiment did not incorporate ^{32}P -NAD on a second exposure to the toxin (not shown).

cells to BFA may involve an altered form of a Golgi G_{α} .

Although there is no evidence at present for a receptor interaction with Golgi G_{α} , one may exist that transfers information from the lumen of the secretory pathway to regulate coat-protein formation on the Golgi. Thus a degree of regulation of the housekeeping function of secretion is possible that was previously unsuspected. □

Received 25 September 1991; accepted 22 January 1992.

- Misumi, Y., Miki, K., Takatsuki, A., Tamura, G. & Ikehara, Y. *J. Biol. Chem.* **261**, 11398–11403 (1986).
- Oda, K. *et al.* *FEBS Lett.* **214**, 135–138 (1987).
- Fujiwara, T., Oda, K., Yokota, S., Takatusuki, A. & Ikehara, Y. *J. Biol. Chem.* **263**, 18545–18552 (1988).
- Doms, R. W., Russ, G. & Yewdell, J. W. *J. Cell Biol.* **109**, 61–72 (1989).
- Lippincott-Schwartz, J., Yaun, L. C., Bonifacio, J. S. & Klausner, R. D. *Cell* **54**, 209–220 (1989).
- Fujiwara, T., Oda, K. & Ikehara, Y. *Cell Struct. Funct.* **14**, 605–616 (1989).
- Ulmer, J. & Palade, G. *Proc. natn. Acad. U.S.A.* **86**, 6992–6996 (1989).
- Lippincott-Schwartz, J. *et al.* *Cell* **60**, 821–836 (1990).
- Duden, R., Griffiths, G., Frank, R., Argos, P. & Kreis, T. E. *Cell* **64**, 649–665 (1991).
- Serafini, T. *et al.* *Nature* **349**, 215–220 (1991).
- Donaldson, J. G., Lippincott-Schwartz, J., Bloom, G. S., Kreis, T. E. & Klausner, R. D. *J. Cell Biol.* **111**, 2295–2306 (1990).
- Donaldson, J. G., Lippincott-Schwartz, J. & Klausner, R. D. *J. Cell Biol.* **112**, 579–578 (1991).
- Gilman, A. G. *Rev. Biochem.* **56**, 615–649 (1987).
- Kahn, R. A. *J. Biol. Chem.* **266**, 25595–25597 (1991).
- Higashijima, T., Uzu, S., Nakajima, T. & Ross, E. M. *J. Biol. Chem.* **263**, 6491–6494 (1988).
- Higashijima, T., Burnier, J. & Ross, E. M. *J. Biol. Chem.* **265**, 14176–14186 (1990).
- Simons, K. & Virta, H. *EMBO J.* **6**, 2241–2247 (1987).
- Ktistakis, N. T., Roth, M. G. & Bloom, G. S. *J. Cell Biol.* **113**, 1009–1023 (1991).
- Orci, L. *et al.* *Cell* **64**, 1183–1195 (1991).
- Donaldson, J. G., Kahn, R. A., Lippincott-Schwartz, J. & Klausner, R. D. *Science* **254**, 1197–1199 (1991).
- Stow, J. L. *et al.* *J. Cell Biol.* **114**, 1113–1124 (1991).
- Balch, W. E., Dunphy, W. G., Braell, W. A. & Rothman, J. E. *Cell* **39**, 405–416 (1984).
- Allan, V. J. & Kreis, T. E. *J. Cell Biol.* **103**, 2229–2239 (1986).

ACKNOWLEDGEMENTS. We thank T. Higashijima for mastoparan analogues and for advice, T. Kreis for monoclonal antibody specific for β -COP, and P. Sternweis for technical assistance; M.E.L. thanks A. G. Gilman for encouragement and research support. This work is dedicated to the memory of D. E. Zwart.

Epstein-Barr virus infection and replication in a human epithelial cell system

Q. X. Li*, L. S. Young*, G. Niedobitek*,
C. W. Dawson*, M. Birkenbach†, F. Wang†
& A. B. Rickinson*‡

*Cancer Research Campaign Laboratories, Department of Cancer Studies, University of Birmingham, Birmingham B15 2TJ, UK

†Brigham & Women's Hospital, 75 Francis Street, Boston, Massachusetts 02115, USA

‡To whom correspondence should be addressed

EPSTEIN-BARR virus, a human herpesvirus with oncogenic potential, infects two target tissues *in vivo*: B lymphocytes, where the infection is largely non-productive¹, and stratified squamous epithelium in which virus replication occurs^{2,3}. The interaction with B cells, initiated through virus binding to the B-cell surface molecule CR2 (ref. 4), has been studied *in vitro* and the virus 'latent' genes associated with B-cell growth transformation defined⁵. By comparison, viral infection of epithelium remains poorly understood, reflecting the lack of an appropriate cell-culture model. Here we describe the development of such a model

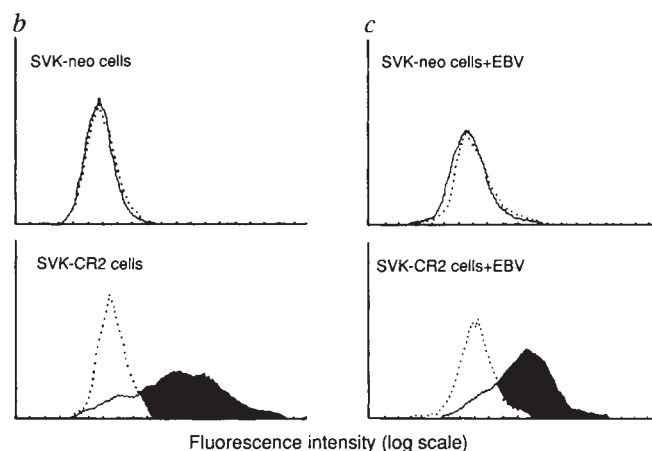
using as targets CR2-expressing transfected cells of two independent human epithelial lines. A high proportion of these cells bind virus and become actively infected, expressing the small EBER RNAs (small non-polyadenylated virus-coded RNAs) and the Epstein-Barr nuclear antigen 1 but not other latent proteins; thereafter, under conditions favouring epithelial differentiation, up to 30% of the cells can be induced to enter virus productive cycle with some progressing to full virus replication. We find significant differences between laboratory virus strains in their ability to infect epithelium that do not correlate with their B-cell growth-transforming activity.

The absence of a demonstrable Epstein-Barr virus (EBV) receptor molecule on available epithelial cell lines, and on most if not all cells in primary keratinocyte culture, is probably the first barrier to the development of a permissive *in vitro* system⁶⁻¹⁰. To overcome this barrier two human epithelial cell lines, SVK and SCC12F, were transfected with a CR2 expression vector, pZip-NeoSV(X)1-CR2, containing a drug-selection marker and the CR2 complementary DNA linked to a retroviral promoter (Fig. 1 legend). High CR2-expressing clones (SVK-CR2, SCC12F-CR2) were selected on the two cell backgrounds, as were control clones (SVK-Neo, SCC12F-Neo) transfected with the pZip-NeoSV(X)1 drug-resistance plasmid¹¹. A pool of anti-CR2 monoclonal antibodies immunoprecipitated the 145K CR2 protein (M_r 145,000) from ¹²⁵I-surface-labelled cells of the SVK-CR2 line and of the human B-cell line Raji, but not from

FIG. 1 Expression of CR2 in SVK transfectants. a,

Immunoprecipitations from surface ¹²⁵I-labelled Raji-BL, SVK-Neo and SVK-CR2 cells using either a pool of the anti-CR2 monoclonal antibodies HB5, OKB7, BU33 and BU35 (lanes 1) or the anti-HLA class I monoclonal antibody W6/32 (lanes 2). Molecular mass markers are shown to the right ($\times 10^{-3}$). b, Cytofluorimetric analysis of surface CR2 expression. Viable cell suspensions of SVK-Neo (top panel) and SVK-CR2 (bottom panel) were exposed to the anti-CR2 monoclonal antibody OKB7 (continuous line), or to the control anti-CD5 monoclonal antibody T1B (dotted line), then stained with fluorescein-labelled goat anti-mouse IgG. SVK-Neo cells showed no CR2 staining (mean fluorescence intensity (m.f.i.) 58, background m.f.i. 59), but SVK-CR2 cells stained strongly (m.f.i. 106; background m.f.i. 67; 71% of the population were in the shaded area above background). In the same assay Raji cells gave an m.f.i. of 100 versus background m.f.i. of 51, with 76% of the population staining above background. Similar results were obtained on 5 successive occasions. c, Cytofluorimetric analysis of EBV binding. Viable cell suspensions of SVK-Neo (top panel) and SVK-CR2 (bottom panel) were incubated with an Akata strain EBV preparation and subsequently exposed to an EBV antibody-positive human serum (continuous line) or to an EBV antibody-negative human serum as a control (dotted line), then stained with fluorescein-labelled goat anti-human IgG. SVK-Neo cells did not bind EBV (m.f.i. 69; background m.f.i. 74), but SVK-CR2 cells stained strongly (m.f.i. 104; background m.f.i. 84; 67% of the population were in the shaded area above background). Similar results were obtained on 4 occasions.

METHODS. Cytoplasmic RNA was extracted from Raji cells by the guanidinium isothiocyanate/caesium chloride procedure. Double-stranded, *Eco*RI methylated cDNA was prepared³⁰ using avian myeloblastomatosis virus reverse transcriptase (Life Sciences) and oligo(dT) primers, and inserted into the *Eco*RI site of bacteriophage λ gt10. The library was screened with ³²P end-labelled oligonucleotide probe synthesized using the sequence of an incomplete cDNA clone³¹. Initial screening resulted in the isolation of several incomplete clones, the largest of which measured 1.76 kilobase pairs. This fragment was used as a probe for subsequent library screening, resulting in the isolation of 3 larger clones, each consisting of multiple *Eco*RI fragments. *Eco*RI restriction fragments were inserted into pBluescript and

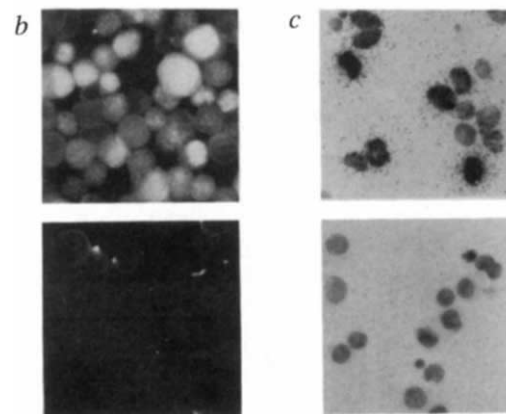
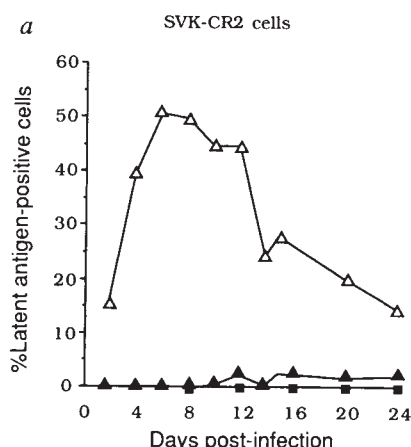


sequenced. Sequences were identical to those already published³¹. A variable exon encoding the short consensus repeat designated number 11 (ref. 32) or 10b (ref. 33) was not present in any of the clones. The full-length CR2 open reading frame was reconstructed by a 3-way ligation of the two protein-encoding *Eco*RI cDNA fragments into the *Eco*RI site of pBluescript. The intact CR2 open reading frame was then inserted into a repaired *Bam*HI site of the expression vector, pZip-NeoSV(X)1 (ref. 11) generating pZip-NeoSV(X)1-CR2. Transfections were carried out using human squamous epithelial cell lines: SVK is an SV40-immortalized, non-tumorigenic skin keratinocyte line; SCC12F, a non-tumorigenic subclone of a facial squamous carcinoma cell line, is free of DNA virus sequences and shows many of the properties of normal human epithelial cells *in vitro*^{28,34}. Exponentially growing cultures of these lines were washed twice with Opti-MEM1 (BRL) and then incubated with a mixture containing 10 μ g plasmid DNA and Lipofectin (BRL). After 4 h, cultures were washed and fed again with normal growth medium (NGM: Jokic's medium, 10% fetal calf serum, 2 mM glutamine and 0.4 μ g ml⁻¹ hydrocortisone for SVK cells, and DME:F12 (3:1) medium with these supplements for SCC12F cells) and left for 3 days. Cells were then collected, replated at 10^4 – 5×10^4 per 10-cm dish and left for 24 h before the addition of 400 μ g ml⁻¹ G418. After 4 weeks of culture, drug-resistant colonies were picked and expanded in G418-containing NGM. Clones transfected with pZip-NeoSV(X)1-CR2 were screened for CR2 expression and highly expressing clones (SVK-CR2, SCC12F-CR2) were selected for experiments, with control clones (SVK-Neo, SCC12F-Neo) transfected with pZip-NeoSV(X)1. Full details of surface iodination, of immunoprecipitation and of the anti-CR2 monoclonal antibodies are given elsewhere¹⁰. Cytofluorimetric analysis of CR2 expression and EBV binding was as described for epithelial cells²⁸.

FIG. 2 EBV latent gene expression in virus-infected SVK-CR2 cells. *a*, Percentage of cells expressing EBNA1 (Δ), EBNA2 ($\blacktriangle$) and LMP1 ($\blacksquare$) up to 24 days postinfection. Representative results from one of 3 experiments. *b*, Anti-complement immunofluorescence staining for EBNA1 in SVK-CR2 cells on day 5 postinfection. Upper panel, staining with a human serum Amo monospecific for EBNA1; lower panel, background staining with an EBV antibody-negative human serum CD. *c*, *In situ* hybridization for EBER RNAs in SVK-CR2 cells on day 6 postinfection. Upper panel, hybridization with antisense EBER1 and EBER2 riboprobes; lower panel, background hybridization with sense EBER1 and EBER2 riboprobes.

METHODS. SVK-CR2 monolayer cultures

at 70–80% confluence were washed twice in PBS and incubated with EBV (filtered supernatant from induced Akata cell cultures) for 4 h at 37 °C, then fed again with NGM. Cultures were re-fed with fresh medium every 2–3 days. Individual cultures were collected by trypsinization at intervals up to 24 days postinfection, the cells washed in PBS and spread on slides as cell smears for immunofluorescence staining or as cytospin preparations for *in*



situ hybridization. Slides were fixed and stained for EBNA1 by anticomplement immunofluorescence using a monospecific human serum Amo (ref. 26), and for EBNA2 and LMP1 by indirect immunofluorescence using antibodies PE2 and CS1-4 respectively^{15,16}. ³⁵S-labelled riboprobes were used for *in situ* hybridizations for EBER RNAs³⁵.

SVK-Neo (lanes 1; lanes 2 show reference immunoprecipitations of the 45K HLA class I heavy chain from the same lines). Immunofluorescence assays on viable cell suspensions confirmed that most SVK-CR2 cells express CR2 on their surfaces at levels equivalent to those observed with the Raji cell line (Fig. 1b), and showed that SVK-CR2 cells could bind EBV (Fig. 1c). The control SVK-Neo-transfected cell line was negative in all assays. Similar results were obtained with the corresponding SCC12F-transfected cell lines (data not shown).

In contrast to CR2-transfected rodent cells (refs 12, 13; our unpublished observations) where expression of the EBV receptor molecule permits virus binding but generally is not followed by active infection, both CR2-positive human epithelial lines are susceptible to the virus. Following exposure to EBV (Akata strain), most cells in SVK-CR2 cultures become infected and express the Epstein-Barr nuclear antigen (EBNA)-1, detectable by immunofluorescence staining with a monospecific human serum (Fig. 2a, b) and by immunoblotting (Fig. 3d). The proportion of EBNA1-positive cells peaked by day 5 postinfection (50–90% in >20 experiments) and significant numbers were still detectable at day 24 (Fig. 2a). A similar pattern was observed when the cultures were screened by *in situ* hybridization for expression of the small non-polyadenylated EBER RNAs (Fig. 2c), another viral marker associated with latent infection¹⁴. But immunofluorescence assays using monoclonal antibodies against other latent proteins^{15,16} revealed no detectable expression of the latent membrane protein-1 (LMP1) and only a few cells, appearing at day 12 postinfection or later, which stained weakly for the nuclear antigen EBNA2 (Fig. 2a). The significance of this staining is unknown. Experimental infection of SCC12F-CR2 cells gave the same pattern of EBV latent gene expression as seen with SVK-CR2, with EBNA1 detectable in up to 70% of cells (data not shown).

We next examined such cultures for evidence of entry into virus productive cycle using monoclonal antibodies against the immediate early protein BZLF1 (ref. 17), the 85K cytoplasmic early antigen¹⁸, and against the late virus capsid antigen¹⁹. Up to 10% of cells in SVK-CR2 and SCC12F-CR2 cultures became BZLF1-positive within a few days of infection, but only a few of these progressed to positivity for the cytoplasmic early antigen or the virus capsid antigen (Fig. 3a). In oral hairy leukoplakia, a focus of EBV replication in epithelial cells *in vivo*, progression through the virus productive cycle depends on ordered squamous epithelial cell differentiation^{3,17}. In our *in vitro* sys-

tem, therefore, we subjected the cells either to nutritional starvation from the time of infection or to treatment with the phorbol ester 12-*O*-tetradecanoylphorbol-13-acetate (TPA, an inducer of EBV replication in B cells) from 4 days postinfection; both treatments elicit a differentiation-like response in SVK-CR2 and in SCC12F-CR2 cells which is detectable by changes in cell morphology and by increased expression of epithelial differentiation markers. Starvation or TPA treatment markedly increased the proportion of SVK-CR2 cells with three such differentiation markers (Fig. 3b, left panel): transglutaminase²⁰, involucrin²¹ and crosslinked envelope formation²². Cultures induced to differentiate show a dramatic increase in the numbers of infected cells entering virus productive cycle by day 7 (Fig. 3b, right panel). Thus 20–30% SVK-CR2 cells became BZLF1-positive

TABLE 1 Biological activity of EBV strains on B cells and SVK-CR2 cells

Virus strain	Virus type*	B cell infectivity† (transforming units ml ⁻¹)	Epithelial cell infectivity‡ (% EBNA 1-positivity)
B95.8	type 1	10 ⁴	4
MCUV5	type 1	10 ³	3
M-ABA	type 1	10 ²	7
AKATA	type 1	10 ²	86
BL74	type 1	0.5 × 10 ²	11
AG876	type 2	8 × 10 ³	29
P3HR1	type 2	non-transforming	5

* EBV strains can be classified as type 1 or type 2 according to polymorphisms in the genes encoding EBNA2, 3A, 3B and 3C (ref. 38).

† Virus preparations from the different producer cell lines were assayed for B cell-transforming activity; a transforming titre of 10⁴ units ml⁻¹ indicates that 50% of indicator B cell cultures exposed to a 1:10⁴ dilution of the virus preparation develop transformed foci. The source of P3HR1 virus, a non-transforming EBNA 2 deletion mutant, was the P3HR1 subclone 13 which is free of heterogeneous DNA (ref. 39).

‡ Undiluted samples of the virus preparations, whose B cell transforming activity had been titrated as above, were assayed on SVK-CR2 cells by measuring the % EBNA 1-positive cells present 5 days postinfection. Values shown are from one of three experiments. Relative differences between the virus strains in their infectivity for SVK-CR2 cells, as measured by EBNA 1 staining, were confirmed by BZLF1 staining at day 7 postinfection, although by comparison absolute numbers of BZLF1-positive cells were always lower.

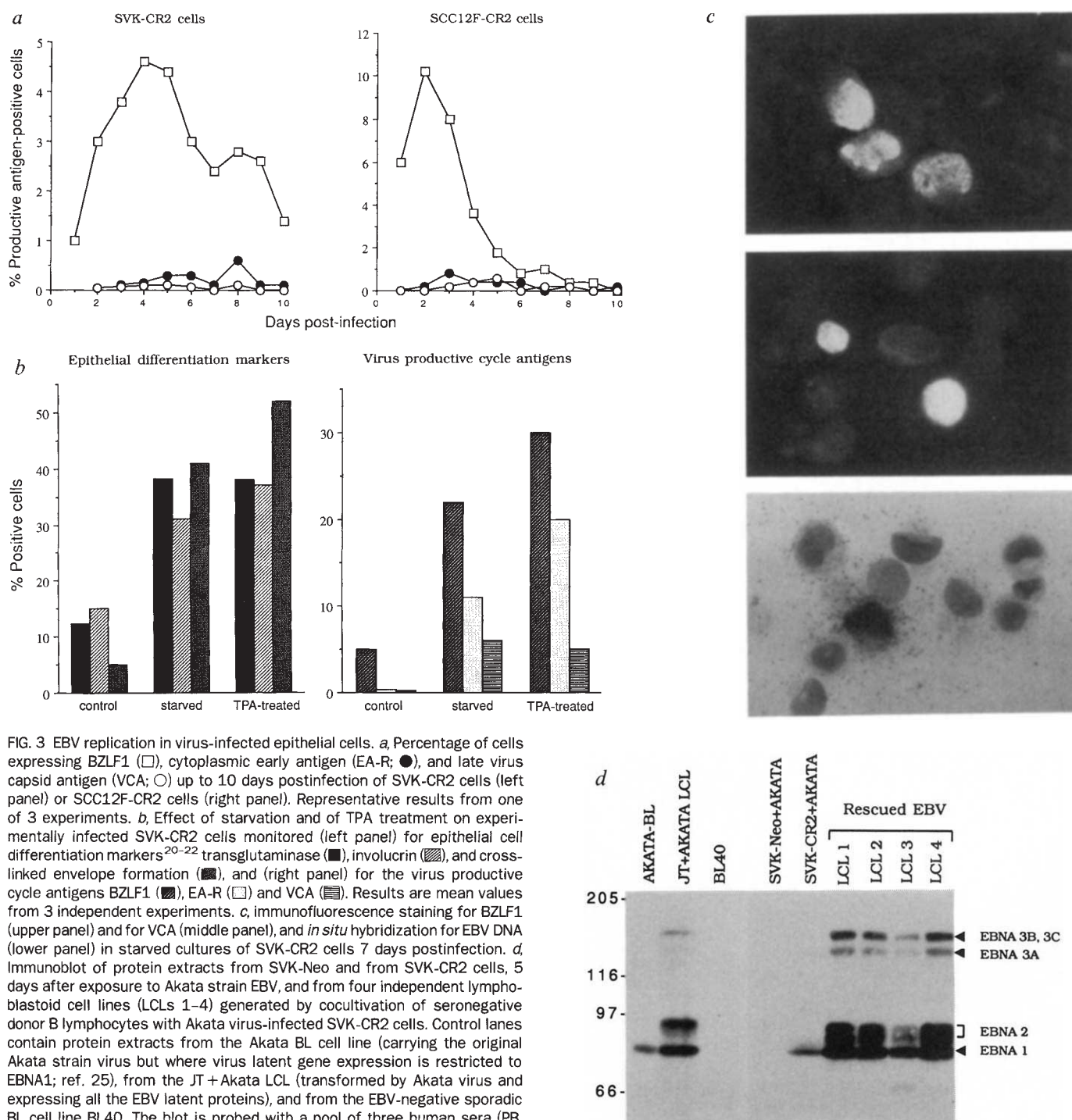


FIG. 3 EBV replication in virus-infected epithelial cells. **a**, Percentage of cells expressing BZLF1 ($\square$), cytoplasmic early antigen (EA-R; $\bullet$), and late virus capsid antigen (VCA; $\circ$) up to 10 days postinfection of SVK-CR2 cells (left panel) or SCC12F-CR2 cells (right panel). Representative results from one of 3 experiments. **b**, Effect of starvation and of TPA treatment on experimentally infected SVK-CR2 cells monitored (left panel) for epithelial cell differentiation markers²⁰⁻²² transglutaminase ($\blacksquare$), involucrin ($\square$), and cross-linked envelope formation ($\blacksquare$), and (right panel) for the virus productive cycle antigens BZLF1 ($\blacksquare$), EA-R ($\square$) and VCA ($\blacksquare$). Results are mean values from 3 independent experiments. **c**, immunofluorescence staining for BZLF1 (upper panel) and for VCA (middle panel), and *in situ* hybridization for EBV DNA (lower panel) in starved cultures of SVK-CR2 cells 7 days postinfection. **d**, Immunoblot of protein extracts from SVK-Neo and from SVK-CR2 cells, 5 days after exposure to Akata strain EBV, and from four independent lymphoblastoid cell lines (LCLs 1-4) generated by cocultivation of seronegative donor B lymphocytes with Akata virus-infected SVK-CR2 cells. Control lanes contain protein extracts from the Akata BL cell line (carrying the original Akata strain virus but where virus latent gene expression is restricted to EBNA1; ref. 25), from the JT+Akata LCL (transformed by Akata virus and expressing all the EBV latent proteins), and from the EBV-negative sporadic BL cell line BL40. The blot is probed with a pool of three human sera (PB, AMo, RS22) containing antibodies against EBNA1, 2, 3A, 3B and 3C, and M_r markers ($\times 10^{-3}$) are shown on the left. Note that the Akata virus-encoded EBNA3B and EBNA3C proteins comigrate under these conditions, whereas the EBNA2 protein usually appears as a doublet; identification of the individual EBNA3s could be confirmed in other gels probed with monospecific sera. Experimentally infected SVK-CR2 cells show selective expression of EBNA1; LCLs generated from cocultures of indicator B cells and experimentally infected SVK-CR2 cells express EBNA1, 2, 3A, 3B and 3C proteins whose sizes are characteristic of the Akata virus strain, confirming the identity of the rescued virus.

METHODS. Cultures were infected with Akata strain virus (Fig. 2) and either re-fed with complete changes of medium every 2-3 days (Fig. 3a and control, Fig. 3b), maintained without medium change (starved, Fig. 3b), or re-fed as the controls but exposed to 5×10^{-8} M TPA from day 4 postinfection (TPA-treated, Fig. 3b). Indirect immunofluorescence staining for BZLF1, EA-R and VCA (Fig. 3a, b, c) was carried out using monoclonal antibodies BZ1, R63 and V3 respectively¹⁷⁻¹⁹, and for involucrin and transglutaminase (Fig. 3b) using a polyclonal rabbit antiserum²⁸ and monoclonal antibody B.CI²⁰

respectively. Crosslinked envelope formation (Fig. 3b) was assayed as described^{22,28} in cells collected on day 7 postinfection from control, starved and TPA-treated cultures and treated overnight with A23187 ionophore. *In situ* hybridization for EBV DNA (Fig. 3c) was carried out with a ³⁵S-labelled BamHI W DNA probe³⁶. For the virus rescue experiments, SVK-CR2 cells and SVK-Neo control cells were exposed to Akata virus, washed and cultured for 7 days without further medium change either in the presence or absence of 100 μ M Acyclovir (ACV), a drug that specifically blocks EBV DNA replication in productive cycle³⁷. The cells were collected, exposed for 3 h to medium containing 25% virus-neutralizing pooled human serum (to eliminate the possibility of any carry-over of residual virus from the original inoculum), washed and cocultivated with indicator B lymphocytes from a seronegative donor in the continued presence or absence of ACV. Lymphoblastoid cell lines were generated from 5/20 cocultures containing SVK-CR2 cells maintained in ACV-free medium postinfection; no transforming virus was rescued in parallel cocultures containing either virus-infected SVK-CR2 cells maintained in ACV medium or virus-exposed SVK-Neo cells.

and 5–6% progressed through to the late phase of the cycle, identifiable by virus capsid antigen staining and by *in situ* hybridization with a radiolabelled *Bam*HI W EBV-DNA probe, when cells replicating viral DNA show an intense signal⁸ (Fig. 3b, c). Two-colour immunofluorescence revealed that a small number of LMP1-positive cells now appeared in the cultures within the productively infected population (data not shown). Progeny virions produced by the epithelial cells could be rescued into indicator B lymphocytes and the EBV-transformed lymphoblastoid cell lines thus generated were shown by EBNA immunoblotting to carry the Akata virus strain (Fig. 3d). Similar results for virus productive antigen expression and for virus rescue were also obtained with experimentally infected SCC12F-CR2 cultures under differentiation-inducing conditions (data not shown).

Finally we compared a number of different laboratory strains of EBV for their infectivity in these epithelial systems, monitoring the appearance both of EBNA1 and of BZLF1-positive cells in infected cultures. There are differences between individual viral strains, consistently observed on both SVK-CR2 and SCC12F-CR2 target cells, which do not correlate with B-cell transforming ability. The most efficient epithelial cell infections have been achieved with the type-1 Akata strain and with the type 2 AG876 strain (Table 1). Further analysis of AG876-infected epithelial cultures reveals a similar sequence of events to those for Akata virus infection (data not shown). The non-transforming EBNA2 deletion mutant P3HR1 shows some activity in epithelial systems whereas the type-1 EBV strain B95.8 gives poor infection, despite using preparations with potent B cell-transforming ability.

Our results reveal key differences in viral behaviour in epithelial versus lymphoid cells, and points to considerable biological heterogeneity amongst virus isolates in their infectivity in the two environments. The initial events of B-cell infection *in vitro* involve expression of EBER RNAs, all 6 EBNA proteins (EBNAs 1, 2, 3A, 3B, 3C, -LP) and LMP1; this leads to cell growth transformation, with only the occasional transformed cell entering virus productive cycle^{5,23}. By contrast, infection in the epithelial cell system seems to involve expression of only the EBER RNAs and EBNA1, followed by a differentiation-linked switch into virus productive cycle. This restricted pattern of EBER RNA/EBNA1 expression, although not identified within oral hairy leukoplakia lesions²⁴, has clear parallels with that seen in Burkitt's lymphoma²⁵ and in the epithelial malignancy nasopharyngeal carcinoma^{24,26,27}. Many cases of nasopharyngeal carcinoma also express LMP1^{26,27}, a protein with the capacity to deregulate epithelial cell growth/differentiation^{28,29}. But in the model systems described here, LMP1 is rarely expressed and then only as part of a cell differentiation-linked productive cycle. If this *in vitro* behaviour in two different human epithelial cell lines reflects the normal situation *in vivo*, then it could be the inappropriate expression of LMP1 in an undifferentiated epithelial cell that underlies the role of the virus in the pathogenesis of nasopharyngeal carcinoma. The natural course of an EBV infection in epithelium is virus replication; the present approach provides a means of studying this differentiation-dependent process *in vitro* and of defining the viral genes that influence the efficiency of infection in this cellular environment. □

10. Young, L. S., Dawson, C. W., Brown, K. W. & Rickinson, A. B. *Int. J. Cancer* **43**, 786–794 (1989).
11. Cepko, C., Roberts, B. & Mulligan, R. *Cell* **37**, 1053–1062 (1984).
12. Ahearn, J. M., Hayward, S. D., Hickey, J. C. & Fearon, D. T. *Proc. natn. Acad. Sci. U.S.A.* **85**, 9307–9311 (1988).
13. Cantaloube, J. F. et al. *Eur. J. Immun.* **20**, 409–416 (1990).
14. Lerner, M. R., Andrews, N. C., Miller, G. & Steitz, J. A. *Proc. natn. Acad. Sci. U.S.A.* **78**, 805–809 (1981).
15. Young, L. et al. *New Engl. J. Med.* **321**, 1080–1085 (1989).
16. Rowe, M. et al. *J. gen. Virol.* **68**, 1575–1586 (1987).
17. Young, L. S. et al. *J. Virol.* **65**, 2868–2874 (1991).
18. Pearson, G. R. et al. *J. Virol.* **47**, 193–201 (1983).
19. Vroman, B., Luka, J., Rodriguez, M. & Pearson, G. R. *J. Virol.* **52**, 107–113 (1985).
20. Thacher, S. M. & Rice, R. H. *Cell* **40**, 685–695 (1985).
21. Rice, R. H. & Green, H. *Cell* **18**, 681–694 (1979).
22. Rubin, A. L. & Rice, R. H. *Cancer Res.* **46**, 2356–2361 (1986).
23. Alfieri, C., Birkenbach, M. & Kieff, E. *Virology* **181**, 595–608 (1991).
24. Gilligan, K., Rajadurai, P., Resnick, L. & Raab-Traub, N. *Proc. natn. Acad. Sci. U.S.A.* **87**, 8790–8794 (1990).
25. Rowe, M. et al. *EMBO J.* **6**, 2743–2751 (1987).
26. Young, L. S. et al. *J. gen. Virol.* **69**, 1051–1065 (1988).
27. Fahraeus, R. et al. *Int. J. Cancer* **42**, 329–338 (1988).
28. Dawson, C. W., Rickinson, A. B. & Young, L. S. *Nature* **344**, 777–780 (1990).
29. Fahraeus, R., Rymo, L., Rhim, J. S. & Klein, G. *Nature* **345**, 447–449 (1990).
30. Gubler, U. & Hoffman, B. J. *Gene* **25**, 263–269 (1983).
31. Weis, J. J. et al. *Proc. natn. Acad. Sci. U.S.A.* **83**, 5639–5643 (1986).
32. Moore, M. D. et al. *Proc. natn. Acad. Sci. U.S.A.* **84**, 9194–9198 (1987).
33. Weis, J. J. et al. *J. exp. Med.* **167**, 1047–1066 (1988).
34. Parkinson, E. K., Pera, M. F., Emmerson, A. & Gorman, P. A. *Carcinogenesis* **5**, 1071–1077 (1984).
35. Milani, S., Herbst, H., Schuppan, D., Hahn, E. G. & Stein, H. *Hepatology* **10**, 84–92 (1989).
36. Niedobitek, G. et al. *Hum. Path.* **20**, 796–799 (1989).
37. Lin, J. et al. *J. Virol.* **50**, 50–55 (1984).
38. Sample, J. et al. *J. Virol.* **64**, 4084–4092 (1990).
39. Heston, L., Rabson, M., Brown, N. & Miller, G. *Nature* **295**, 160–163 (1982).

ACKNOWLEDGEMENTS. We thank S. Williams for photography and D. Williams for secretarial help. This work was supported by the Cancer Research Campaign; G.N. was supported by the Deutsche Forschungsgemeinschaft.

Quantal calcium release by purified reconstituted inositol 1,4,5-trisphosphate receptors

Christopher D. Ferris*, Andrew M. Cameron*, Richard L. Huganir*† & Solomon H. Snyder*‡§||

* Departments of Neuroscience, ‡ Pharmacology and Molecular Sciences, § Psychiatry and Behavioral Sciences, † The Howard Hughes Medical Institute, Johns Hopkins University School of Medicine, 725 North Wolfe Street, Baltimore, Maryland 21205, USA

RELEASE of intracellular Ca^{2+} by inositol 1,4,5-trisphosphate (InsP_3) occurs through specific receptor proteins¹ which are ligand-activated Ca^{2+} channels². Changes in intracellular Ca^{2+} regulate many cellular functions³. This Ca^{2+} release is a discontinuous quantal process in which successive increments of InsP_3 transiently release precise amounts of Ca^{2+} (refs 4–6). Possible explanations of quantal Ca^{2+} release have included rapid degradation of InsP_3 , reciprocity of Ca^{2+} release and sequestration, desensitization of InsP_3 receptors⁷, or actions of InsP_3 on discrete compartments of Ca^{2+} with variable sensitivity to InsP_3 (ref. 4). We successfully reconstituted InsP_3 -induced Ca^{2+} flux in vesicles containing only purified InsP_3 receptor protein². The reconstituted vesicles retain the regulatory features of the InsP_3 receptor, including phosphorylation sites⁸ and modulation of Ca^{2+} release by adenine nucleotides⁹. Using these reconstituted vesicles, we show here that quantal flux of Ca^{2+} elicited by InsP_3 is a fundamental property of its receptor.

We monitored InsP_3 -induced Ca^{2+} flux in reconstituted vesicles as before². In this system, InsP_3 stimulates the rapid equilibration of $^{45}\text{Ca}^{2+}$ into the InsP_3 receptor (InsP_3R)-containing vesicles in the absence of a chemical or voltage gradient. Exposure to different concentrations of InsP_3 triggers an augmentation of $^{45}\text{Ca}^{2+}$ flux which is rapid for the first 10 seconds and then slows (Fig. 1). The level at which $^{45}\text{Ca}^{2+}$ flux reaches a plateau is proportional to the amount of InsP_3 added,

Received 12 September 1991; accepted 20 February 1992.

1. Nilsson, K., Klein, G., Henle, W. & Henle, G. *Int. J. Cancer* **8**, 443–450 (1971).
2. Sixbey, J. W., Nedrud, J. G., Raab-Traub, N., Hanes, R. A. & Pagano, J. S. *New Engl. J. Med.* **310**, 1225–1230 (1984).
3. Greenspan, J. S. et al. *New Engl. J. Med.* **313**, 1564–1571 (1985).
4. Fingerroth, J. D. et al. *Proc. natn. Acad. Sci. U.S.A.* **81**, 4510–4514 (1984).
5. Kieff, E. & Liebowitz, D. in *Virology* (eds Fields, B. N. et al.) 1889–1920 (Raven, New York, 1990).
6. Miller, G., Grogan, E., Heston, L., Robinson, J. & Smith, D. *Science* **212**, 452–454 (1981).
7. Shapiro, I. M. & Volsky, D. J. *Science* **215**, 1225–1228 (1983).
8. Sixbey, J. W. et al. *Nature* **306**, 480–483 (1983).
9. Sixbey, J. W. et al. *J. gen. Virol.* **68**, 805–811 (1987).

|| To whom correspondence should be addressed.

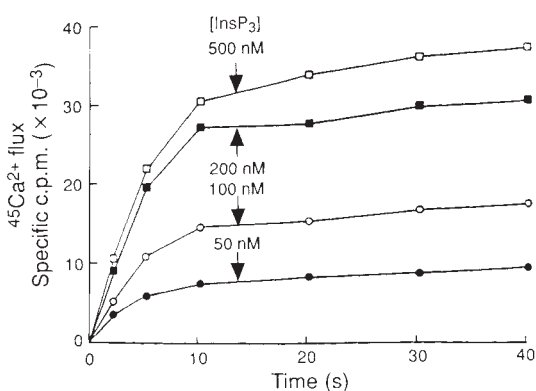


FIG. 1 Submaximal concentrations of InsP_3 activate only a fraction of InsP_3R . Various concentrations of InsP_3 were incubated with reconstituted InsP_3R for the times indicated. Specific c.p.m. refer to the difference between the $^{45}\text{Ca}^{2+}$ flux in the presence and absence of InsP_3 . This difference varied between 6- and 12-fold for individual reconstitution preparations. InsP_3 activates a rapid and biphasic equilibration of $^{45}\text{Ca}^{2+}$. Although in some systems InsP_3 activates Ca^{2+} release on a subsecond timescale²⁴, the time for complete mixing of vesicles with InsP_3 results in an apparent delay such that maximal release is not observed until about 10 s. The slow InsP_3 -dependent phase reaches equilibrium between 10 and 40 s. After 40 s this slow phase appears to stop as no further increase in specific c.p.m. is detected. In a series of experiments ($N=10$) the rate of this slow phase varied between one-tenth and one-fourth of the fast phase. No such slow phase was observed in reconstituted nAChR, suggesting that this slow phase of equilibration may be a specific property of the InsP_3R . As described in the text, 50 nM InsP_3 (filled circles) activates about one-half as much $^{45}\text{Ca}^{2+}$ flux as 100 nM InsP_3 (circles), which is about one-half that of 200 nM InsP_3 (filled squares). Thus, submaximal concentration of InsP_3 activates the rapid

so total flux at 100 nM InsP_3 is about twice what it is at 50 nM, and at 200 nM, but at 500 nM InsP_3 it is only slightly greater than at 200 nM InsP_3 . After 40 s, InsP_3 -stimulated $^{45}\text{Ca}^{2+}$ flux stops, so that final levels at all concentrations of InsP_3 are about the same at 5 min as at 40 s (data not shown).

Receptor desensitization affords one explanation for submaximal $^{45}\text{Ca}^{2+}$ flux at submaximal InsP_3 levels. To investigate this possibility, we exposed vesicles to 50 nM InsP_3 for 10 s then added 500 nM InsP_3 (Fig. 2). No desensitization is apparent, as Ca^{2+} flux elicited by the second application of 500 nM InsP_3 is the same as the response to 500 nM InsP_3 applied to vesicles not previously exposed to InsP_3 . There is no desensitization when vesicles are preincubated with concentrations of InsP_3 in the range 10 nM–1 μM for 5 s–5 min before $^{45}\text{Ca}^{2+}$ is added (data not shown). For comparison, we examined $^{45}\text{Ca}^{2+}$ flux in purified and reconstituted nicotinic acetylcholine receptors (nAChR) (Fig. 2b). Exposure of reconstituted nAChR to 10 μM carbachol activates a submaximal degree of Ca^{2+} flux and 200 μM activates a nearly maximal increase in flux. But when 200 μM carbachol is applied 20 s after 10 μM carbachol, there is no additional increase in influx (Fig. 2b), reflecting rapid desensitization of the nAChR as previously reported^{10,11}.

Successive addition of increasing amounts of InsP_3 to permeabilized basophils elicits arithmetic increases in Ca^{2+} release, providing a precise signal detection system⁶. To determine whether this is an inherent property of InsP_3 receptors, we exposed reconstituted vesicles to successive additions of InsP_3 from 50 to 200 nM (Fig. 3). Two successive exposures to 50 nM InsP_3 give exactly the same augmentation in Ca^{2+} flux as a single exposure to 100 nM. Successive exposures to 50, 50 and 100 nM InsP_3 increase Ca^{2+} flux to the same extent as a single exposure to 200 nM InsP_3 (Fig. 3). Thus, purified InsP_3R , like intact cells, accurately detects incremental increases of InsP_3 .

Our principal finding is that the quantal release of internal Ca^{2+} stores, previously described in intact cells^{4,6}, reflects a

equilibration of only a fraction of the total InsP_3 accessible intravesicular volume. Although much of the response to InsP_3 in this preparation occurs over a small concentration range (20–200 nM), responses are not maximal until micromolar levels. Thus, InsP_3 -activated Ca^{2+} flux may be positively cooperative at low concentrations in this system. Maximal concentrations of InsP_3 can equilibrate about half of the total internal volume as assessed by passive equilibration of $^{45}\text{Ca}^{2+}$ or the Ca^{2+} ionophore A23187 (Molecular Probes). Metabolism of $[^3\text{H}]\text{InsP}_3$ was measured by cation-exchange chromatography. We were unable to detect any InsP_3 metabolism in these reconstituted vesicles under assay conditions, suggesting that submaximal equilibration does not result from loss of InsP_3 . As on average there is less than one receptor per vesicle in these preparations, submaximal doses of InsP_3 activate only a fraction of InsP_3R . As previously reported², the final free Ca^{2+} concentration in these preparations is 500 nM. The data shown are the averages of duplicate determinations, which varied by less than 10% ($n=4$).

METHODS. The time course of InsP_3 -activated $^{45}\text{Ca}^{2+}$ flux for different InsP_3 concentrations was measured as before². Briefly, InsP_3R was isolated from rat cerebellum using heparin-agarose and concanavalin A-Sepharose^{2,13}. Purity of receptor preparations was confirmed by SDS-PAGE and silver staining. After rapid isolation, InsP_3R was reconstituted into phosphatidylcholine/phosphatidylserine vesicles by CHAPS dialysis against buffer A (20 mM Tris, pH 7.4 at 25 °C, 100 mM NaCl, 100 mM KCl, 2 mM β -mercaptoethanol) for 72 h. These vesicles were used to measure either $[^3\text{H}]\text{InsP}_3$ binding or InsP_3 -activated $^{45}\text{Ca}^{2+}$ flux². Flux reactions were initiated by addition of 70 μl reconstituted proteoliposomes to 30 μl buffer A containing 2 μCi $^{45}\text{Ca}^{2+}$ in the presence or absence of InsP_3 . The $^{45}\text{Ca}^{2+}$ flux reaction was stopped by addition of buffer B (buffer A plus 0.5 mM CaCl_2 , 5 mM MgSO_4 , and 100 $\mu\text{g ml}^{-1}$ heparin) and extravesicular $^{45}\text{Ca}^{2+}$ was rapidly removed by cation-exchange chromatography on Dowex (50 WX; Sigma)². Vesicles were collected and intravesicular $^{45}\text{Ca}^{2+}$ counted by liquid scintillation spectrometry. In some experiments nAChR was isolated, reconstituted¹⁰, and proteoliposomes used to measure carbachol-stimulated $^{45}\text{Ca}^{2+}$ flux as for InsP_3R . The use of the term flux refers to the rapid equilibration of $^{45}\text{Ca}^{2+}$ in the presence of InsP_3 or carbachol, rather than to the absolute amount of Ca^{2+} translocated per unit time per membrane area.

fundamental and possibly unique property of InsP_3R . As the purified receptor protein preparation does not contain InsP_3 -metabolizing enzymes or $(\text{Ca}^{2+})\text{ATPases}$, these enzymes must not be required for quantal Ca^{2+} release. Our data also show that desensitization is not necessary for quantal Ca^{2+} release. With the pure receptor reconstituted into vesicles, successive applications of InsP_3 produce increments of Ca^{2+} flux in arithmetic proportion and with precision. By contrast, desensitization follows application of carbachol to reconstituted nAChR. Quantal Ca^{2+} release has been explained by postulating that the sensitivity of the receptor to InsP_3 varies with the Ca^{2+} content of individual Ca^{2+} stores¹². But in our reconstitution system, the concentration of Ca^{2+} inside and outside the vesicles remains constant.

Demonstration that quantal release of Ca^{2+} is a property of isolated InsP_3R limits the possible mechanisms behind this phenomenon. One explanation involves heterogeneity of InsP_3R in the release of Ca^{2+} in response to InsP_3 . According to this model, low concentrations of InsP_3 would enhance Ca^{2+} flux selectively from a more sensitive population of receptors, whereas progressively increasing concentrations of InsP_3 would influence the less sensitive receptors. Such heterogeneity could be secured by variations in receptor affinity for binding InsP_3 . We have tested the binding of $[^3\text{H}]\text{InsP}_3$ to reconstituted InsP_3R . As previously reported², in these preparations we observe a monophasic Scatchard plot ($K_d=30$ nM), as with unreconstituted pure InsP_3R protein¹³, reflecting a single population of binding sites with a Hill coefficient of 1.0 (data not shown). Alternatively, heterogeneity of receptors may lie in the coupling between the InsP_3 recognition site at the N-terminal portion of the receptor protein and the Ca^{2+} channel at the C terminal^{14–17}. Interestingly, molecular cloning reveals multiple forms of InsP_3R derived by alternative splicing, with most of the alternative forms involving variations in this coupling domain of the receptor^{18,19}. Variable sensitivity of InsP_3 -sensitive Ca^{2+} stores to InsP_3 occurs in permeabilized hepatocytes with spontaneous

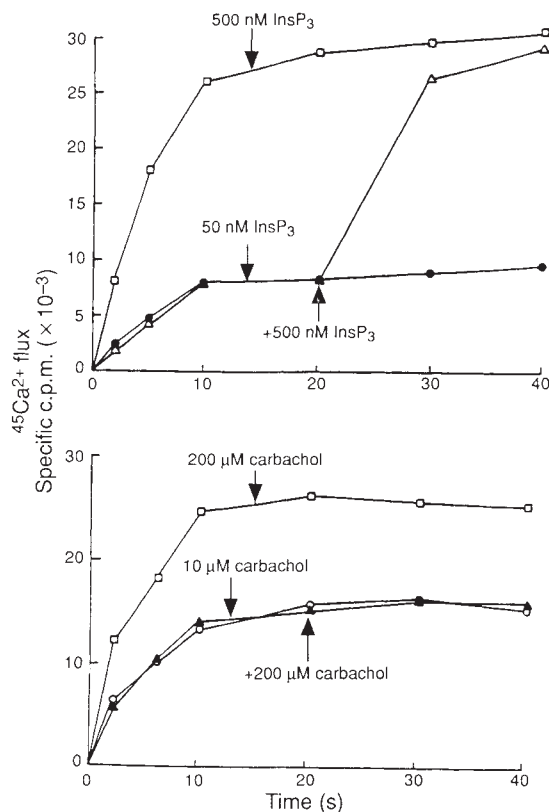


FIG. 2 Desensitization with nAChR but not InsP_3R . **a**, Reconstituted InsP_3R is used for $^{45}\text{Ca}^{2+}$ flux measurement (see legend to Fig. 1). Reconstituted InsP_3R was incubated in the presence of 50 (filled circles) or 500 nM InsP_3 (squares) for the indicated times. In some samples, after 10-s incubation with 50 nM InsP_3 , InsP_3 from a $50\times$ sterile solution was added to increase the final concentration of InsP_3 to 500 nM (triangles) to determine whether InsP_3R is desensitized to InsP_3 . In other experiments, reconstituted InsP_3R was preincubated with InsP_3 (10 nM–1 μM) for times from 5–300 s, followed by addition of $2\text{ }\mu\text{Ci}$ $^{45}\text{Ca}^{2+}$ for 10 s to monitor possible desensitization of InsP_3R to InsP_3 is observed under any of these conditions. For comparison (**b**) reconstituted nAChR is incubated with 10 μM (circles) or 200 μM (squares) carbachol for the indicated times. In some samples, after 10- or 20-s incubation with 10 μM carbachol, 200 μM carbachol was added (filled triangles). nAChR is desensitized, as addition of a near-maximal dose (200 μM) of carbachol following brief incubation with a submaximal dose (10 μM), fails to stimulate further $^{45}\text{Ca}^{2+}$ flux. $^{45}\text{Ca}^{2+}$ flux using reconstituted nAChR was measured as described in the legend to Fig. 1. Data are averages of triplicate determinations which varied by less than 10% ($n=3$).

Ca^{2+} release from portions of the InsP_3 -sensitive stores²⁰. The tetrameric nature of the InsP_3R (refs 13, 21), the existence of several gene products^{22,23}, alternative splicing of messenger^{14,18,19} and other post-translational modifications such as phosphorylation⁸ may generate considerable molecular diversity for InsP_3R . Although such molecular heterogeneity may contribute to quantal Ca^{2+} release, we have not yet been able to develop a quantitative model to simulate our experimental findings.

The quantal release properties of the InsP_3R seem to be unique and are not evident with other plasma membrane ligand or

voltage-gated ion channel proteins. We did not observe such behaviour with reconstituted nAChR and are not aware of any other channel proteins with quantal release properties. The increment detection function of InsP_3R provides a precise system for setting intracellular Ca^{2+} levels at specific values to communicate information with temporal and spatial specificity from intercellular signalling to intracellular Ca^{2+} -dependent regulatory proteins. The precise regulation of Ca^{2+} flow through the InsP_3R and cytosolic Ca^{2+} levels contrasts with ligand and voltage-gated ion channels at the plasma membrane that control and respond to the electrical status of the cell (membrane voltage) and do not regulate the exact cytosolic concentration of any ion that permeates the channel. □

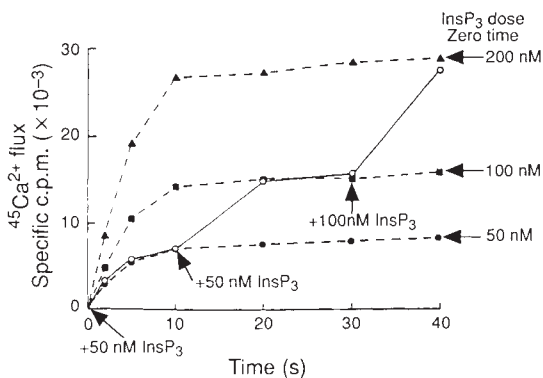


FIG. 3 Quantal Ca^{2+} flux by reconstituted InsP_3R quantitatively detects increments of InsP_3 . InsP_3 -activated $^{45}\text{Ca}^{2+}$ flux was measured as described in the legend to Fig. 1. Filled symbols indicate the time course of $^{45}\text{Ca}^{2+}$ flux at the indicated InsP_3 doses added at time zero. Circles reflect $^{45}\text{Ca}^{2+}$ flux activated by successive additions of InsP_3 (making the new concentration first 100 nM InsP_3 and then 200 nM as indicated by the arrows) to vesicles following 10-s incubation with 50 nM InsP_3 , which enhances $^{45}\text{Ca}^{2+}$ flux to the same extent as 100 nM and then 200 nM InsP_3 . Likewise, successive 10-s incubations with final InsP_3 concentrations of 50 nM, 100 nM and 200 nM quantitatively detect changes in InsP_3 concentrations. Longer incubation with the final addition of 100 nM InsP_3 does not result in any further $^{45}\text{Ca}^{2+}$ equilibration (data not shown). Data are means of duplicate determinations, which varied less than 10% ($n=3$).

Received 10 September 1991; accepted 15 January 1992.

- Ferris, C. D. & Snyder, S. H. *A. Rev. Physiol.* **54**, 469–488 (1991).
- Ferris, C. D., Huganir, R. L., Supattapone, S. & Snyder, S. H. *Nature* **342**, 87–89 (1989).
- Berridge, M. J. & Irvine, R. F. *Nature* **341**, 197–204 (1989).
- Muallem, S., Pandolfi, S. J. & Beeker, T. G. *J. Biol. Chem.* **264**, 205–212 (1989).
- Parker, I. & Inorrra, I. *Science* **250**, 977–979 (1991).
- Meyer, T. & Stryer, L. *Proc. natn. Acad. Sci. U.S.A.* **87**, 3841–3845 (1990).
- Champeil, P. et al. *J. Biol. Chem.* **264**, 17665–17673 (1989).
- Ferris, C. D., Huganir, R. L., Bredt, D. S., Cameron, A. M. & Snyder, S. H. *Proc. natn. Acad. Sci. U.S.A.* **88**, 2232–2235 (1991).
- Ferris, C. D., Huganir, R. L. & Snyder, S. H. *Proc. natn. Acad. Sci. U.S.A.* **87**, 2147–2151 (1990).
- Huganir, R. L. & Racker, E. *J. Biol. Chem.* **257**, 9372–9378 (1982).
- Hopfield, J. F., Tank, D. W., Greengard, P. & Huganir, R. L. *Nature* **336**, 677–679 (1988).
- Irvine, R. F. *FEBS Lett.* **263**, 5–9 (1990).
- Supattapone, S., Worley, P. F., Baraban, J. M. & Snyder, S. H. *J. Biol. Chem.* **263**, 1530–1534 (1988).
- Mignery, G. A., Newton, C. L., Archer, B. T. III & Sudhof, T. C. *J. Biol. Chem.* **265**, 12679–12685 (1990).
- Mignery, G. A. & Sudhof, T. C. *EMBO J.* **9**, 3893–3898 (1990).
- Furuichi, T. et al. *Nature* **342**, 32–38 (1989).
- Miyawaki, A. et al. *Proc. natn. Acad. Sci. U.S.A.* **88**, 4911–4915 (1991).
- Danoff, S. K. et al. *Proc. natn. Acad. Sci. U.S.A.* **88**, 2951–2955 (1991).
- Nakagawa, T., Okano, H., Furuichi, T., Aruga, J. & Mikoshiba, K. *Proc. natn. Acad. Sci. U.S.A.* **88**, 6244–6248 (1991).
- Missiaen, L., Taylor, C. W. & Berridge, M. J. *Nature* **352**, 241–244 (1991).
- Maeda, N. et al. *J. Biol. Chem.* **266**, 1109–1116 (1991).
- Sudhof, T. C., Newton, C. L., Archer, B. T. III, Ushkaryov, Y. A. & Mignery, G. A. *EMBO J.* **10**, 3199–3206 (1991).
- Ross, C. A. et al. *Soc. Neurosci. Abstr.* (1991).
- Meyer, T., Wensel, T. & Stryer, L. *Biochemistry* **29**, 32–37 (1990).

ACKNOWLEDGEMENTS. We thank G. Yellen for discussion and L. Moritz for technical assistance. Supported by USPHS (S.H.S. and C.D.F.), and the Howard Hughes Medical Institute, the International Life Sciences Institute and Bristol-Myers-Squibb (R.L.H.).

The *wee1* protein kinase is required for radiation-induced mitotic delay

Roy Rowley*, James Hudson† & Paul G. Young†

* Department of Radiology, University of Utah Medical Center, Salt Lake City, Utah 84132, USA

† Department of Biology, Queens University, Kingston, Ontario K7L 3N6, Canada

CELLULAR feedback or 'checkpoint' mechanisms maintain the order of completion of essential, cell-cycle related functions¹⁻³. In the budding yeast, for example, the *RAD9* gene product is required to delay progression into mitosis in response to DNA damage^{2,4-6}. Similarly, in fission yeast, the *cdc25* and *cdc2* gene products influence the ability of cells to delay mitosis in response to the inhibition of DNA synthesis⁷. Because these two checkpoint controls regulate the same event, mitosis, we observed the effect of γ -irradiation on cell cycle progression in fission yeast, to test whether the two controls require the same cell-cycle regulatory elements. We show that γ -radiation-induced mitotic delay requires functional *wee1* protein kinase but does not seem to involve the *cdc25* pathway. Mitotic delay in response to DNA damage is thus distinct from the delay induced by inhibition of DNA synthesis, which involves *cdc25* but is not dependent on *wee1*.

In the fission yeast *Schizosaccharomyces pombe*, progression from G2 into mitosis depends on the activity of the protein kinase *cdc2*, the product of the *cdc2* gene^{8,9}. Activation of *cdc2* and entry into mitosis normally require the protein tyrosine phosphatase *cdc25* (refs 10-15). Inactivation of *cdc2* requires the activity of the protein tyrosine kinase *wee1* (refs 16, 17) and the kinase *cdrl1/nim1* (refs 18, 19), which directly or indirectly inactivates *wee1*. This network of controls is described fully elsewhere¹⁸. Inactivation of *cdc2* may also require another kinase, *mik1*, which functions cooperatively with *wee1* (ref. 20). The protein kinase *cdc2* has at least two phosphorylation sites in *S. pombe*, on serine, threonine and tyrosine residues²¹. Dephosphorylation of Tyr 15 occurs as cells progress from G2 into mitosis and artificially induced dephosphorylation of Tyr 15 causes premature entry into mitosis^{12,15}. Extensive genetic analysis of these pathways (see ref. 22 for review) has led to the ability to manipulate the mitotic control through loss-of-function and gain-of-function mutations and by plasmid overexpression. This genetic framework allowed us to test which of the known elements of mitotic control are required for radiation-induced mitotic delay, by monitoring mitotic progression after γ -irradiation. The results are contrasted with the effects of hydroxyurea treatment, which has been used to determine the genetic requirements for cell cycle delay after inhibition of DNA synthesis⁷. The primary actions of these two agents differ: hydroxyurea blocks cell cycle progression by inhibition of DNA synthesis as it inhibits ribonucleotide reductase²³, whereas mitotic delay caused by ionizing radiation is probably a result of DNA damage^{4,24}. Ionizing radiation can also cause a delay in the completion of DNA synthesis, but at higher doses than those required to cause a delay in progression from G2 to mitosis²⁵.

In wild-type *S. pombe*, as in mammalian cells²⁶, radiation-induced mitotic delay is evident as a transient depression in the rate at which cells enter mitosis and divide (Fig. 1a). The timing suggests that mitotic delay results from arrest at or before the G2/M boundary²⁷. Irradiated yeast populations also show a delay in population growth (Fig. 2a). These results were used to characterize the effects of irradiation on cell cycle progression in strains mutant for components of mitotic control. It was previously suggested⁷ that *cdc25* regulates progression into mitosis in response to the inhibition of DNA synthesis and that a *wee1* gain-of-function allele of *cdc2*, *cdc2-3w*, which is relatively

insensitive to *cdc25* function¹⁶, is defective for this control. We therefore tested whether *cdc2-3w* cells are delayed in progression into mitosis in the presence of DNA damage induced by γ -irradiation. Surprisingly, *cdc2-3w* cells are delayed, and their response is indistinguishable from that of wild-type cells (Figs 1b and 2a). Further examination of strains with mutations affecting mitotic control showed that only inactivation of *wee1* prevented radiation-induced mitotic delay. The response of the *wee1-50* loss-of-function strain was measured at permissive

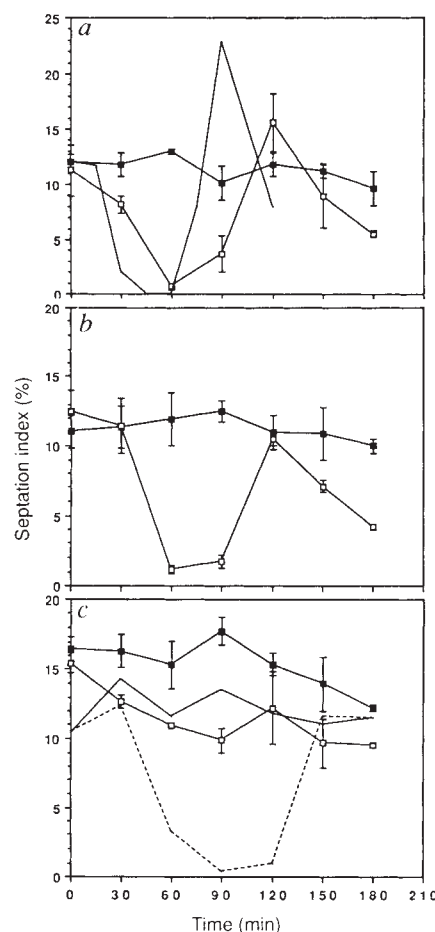


FIG. 1 a, Septation indices (the percentage of cells with forming septa) for strain 972h⁻ (wild type), measured at intervals after irradiation (□, 50 Gy γ -rays) or mock irradiation (■) at 30 °C. Points are means of three estimates, bars indicate the s.e.m. Septum formation begins at the onset of cell division and is considered complete when the daughter cells start to separate. The septation index thus provides a measure of cell division activity in a growing cell population. Following irradiation the index dropped almost to zero (50% of control at 35 min) then recovered (50% of control at 80 min), giving a delay of 45 min at this dose. The time at which the index fell indicates that a block to progression (the γ -ray transition point) was imposed not later in the cell cycle than 35 min before visible septation, or coincident with the G2/M boundary. In wild-type cells maintained at 36 °C (line, no symbols) the septation index fell sooner after irradiation (50% at 25 min) and recovered sooner (50% at 70 min). b, Septation indices for the strain *cdc2-3w*, non-irradiated cells (■) and irradiated cells (□). Points are means for three experiments, bars indicate the s.e.m. c, Septation indices for *wee1-50* cells measured at intervals after γ -irradiation (50 Gy) or mock irradiation. Non-irradiated cells maintained at 25 °C (solid line); irradiated cells maintained at 25 °C (dashed line); non-irradiated cells maintained at 36 °C (■); non-irradiated cells maintained at 36 °C (□).

METHODS. Cells were grown at 25 °C or 30 °C in yeast extract and glucose supplemented with adenine (YEA; ref. 32). Septation was monitored in cells fixed and stained in iodine and observed microscopically at $\times 400$ magnification. Irradiations were performed in a Shepherd caesium irradiator, which gave a maximum dose rate of 8.90 Gy min⁻¹. Cells were irradiated in suspension at a density of $\sim 5 \times 10^6$ cells per ml.

(25 °C) and restrictive (36 °C) temperatures by monitoring septation in an asynchronous population (Fig. 1c), by measurement of population growth delay (Fig. 2b) and by monitoring septation in synchronous cultures (Fig. 3a). For comparison, the response at 36 °C of cells wild-type for *wee1* is shown in Figs 1a, 2a and 3b. No other mutation gave such complete

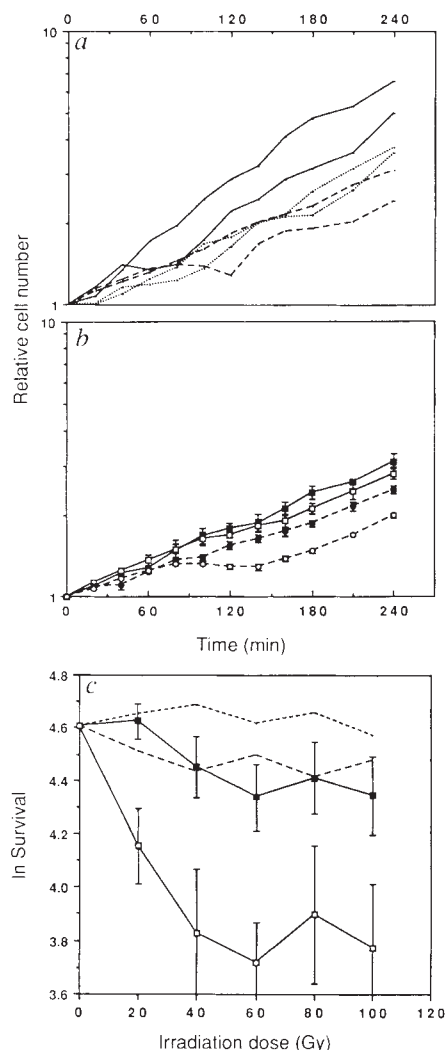


FIG. 2 a, Population growth curves for the wild-type strain *972h⁻* maintained at 30 °C (solid lines), *972h⁻* maintained at 36 °C (dotted lines) and the mutant *cdc2-3w* (dashed lines). The lower line of each pair is for irradiated cells (50 Gy). Cell numbers were estimated by electronic particle counter at intervals after γ -irradiation or mock-irradiation and normalized to the zero-time value. b, Population growth curves for *wee1-50*. Solid lines, cells maintained at 36 °C. Dashed lines, cells irradiated (50 Gy) and maintained at 25 °C; solid symbols, mock-irradiated cells; open symbols, irradiated cells (50 Gy). Points are means of at least three estimates, bars indicate the s.e.m. c, Radiation dose-response curves for *972h⁻* cells maintained at 25 °C (short-dashed line); *972h⁻* cells maintained at 36 °C (long-dashed line); *wee1-50* cells maintained at 25 °C (■); *wee1-50* cells maintained at 36 °C (□); bars indicate s.e.m.

METHODS. To measure survival, cells were grown in suspension at 25 °C, sampled, sonicated, counted and plated on YEA. For each cell line, platings were made from the same cell suspension to minimize variations in the numbers of cells plated. Cells were irradiated on plates (caesium γ -rays, dose rate, 8.90 Gy min⁻¹) then half of the plates for each dose were incubated at 25 °C and half at 36 °C. After 24 h, plates held at 36 °C were moved to a 25 °C incubator. Colonies were scored at 72 h after plating and the per cent plating efficiency calculated as the number of colonies formed divided by the number of cells plated. The plating efficiencies of irradiated cells were then normalized to the plating efficiencies of the non-irradiated samples to give survival percentages. The log_e of the survival percentage is plotted against irradiation dose.

suppression of radiation-induced delay (Table 1). Loss-of-function mutations in *cdr1/nim1* or *cdr2* did not affect mitotic delay induced by radiation. Therefore if the *wee1* product must be activated after irradiation to impose G2 delay, it is not through inactivation of the *cdr1/nim1* kinase, perhaps implying that the signal for delay requires activation of a competing phosphatase instead. As *wee1* and *mik1* kinase activities are sufficient to maintain phosphorylation of *cdc2* on Tyr 15 (ref. 20), the effect of a deletion mutation of *mik1* (*mik1::ura4*) on mitotic-delay induction was examined. The result clearly distinguished between the two kinases: *wee1* is required for radiation-induced delay, *mik1* is not (Table 1). Because of the central role of *cdc2* Tyr-15 phosphorylation in the regulation of G2 progression^{12,15}, further experiments were done to establish whether conditions that reduce *cdc2* Tyr-15 phosphorylation might abolish the radiation effect. Indeed overexpression of wild-type *cdc25* phosphatase and expression of a human protein tyrosine phosphatase under strong promotion¹² both reduced the radiation-induced delay (Table 1). The *stf1-1* mutation, which provides a dominant bypass of *cdc25* function²⁷ did not alter the wild-type radiation response.

If radiation-induced mitotic delay provides time for the cell to repair potentially lethal DNA damage before entry into mitosis, the temperature-sensitive *wee1-50* mutant strain should be sensitive to radiation at 36 °C. Figure 2c shows cell survival rates at different radiation doses for *wee1-50* cells either maintained at 25 °C, or irradiated at 25 °C and kept at 36 °C for 24 h after irradiation. Inactivation of *wee1* after irradiation does render the cell sensitive to radiation, although this effect was only observed at doses under 200 Gy. The effect is small compared with the response of the budding yeast checkpoint mutant, *rad9* (ref. 4), perhaps because radiation doses of ≥ 250 Gy do transiently suppress septation activity in *wee1-50* (36 °C), although less than in wild type. Delays at high doses may be due to other forms of cell cycle perturbation²⁵ or possibly to suppression of the mutant defect by a stronger (dose-dependent) signal for *wee1* activation and progression arrest. As an additional demonstration that inactivation of *wee1* modifies the cellular response to DNA damage, a double mutant *wee1-50cdc17-K42* was made and the loss of viability at the restrictive temperature compared to the loss of viability of *cdc17-K42*. The single mutant, *cdc17-K42* bears a temperature-sensitive conditional mutation of DNA ligase. Inactivation of DNA ligase blocks cell cycle progression in late S or G2 and newly synthesized DNA remains at the size of Okazaki fragments²⁹. The

TABLE 1 Summary of responses to ionizing radiation or hydroxyurea of various fission yeast strains

Strain	Radiation-induced delay?	Hydroxyurea-induced delay?
972 (wild type)	yes	yes
<i>cdc2-3w</i>	yes	no ⁷
<i>wee1-50</i> (25 °C)	yes	yes ⁷
<i>wee1-50</i> (36 °C)	no	yes ⁷
<i>mik1::ura4.D18</i>	yes	partial†
<i>leu1-32 pSAM1.46(nim1 ADH)</i>	reduced	ND
<i>cdr1-76</i>	yes	yes
<i>cdc25-22(cdc25 ADH)</i>	reduced	no ⁷
T-cell phosphatase*	reduced	ND
<i>cdr2-96</i>	yes	yes
<i>cdc2-1w</i>	yes	yes ⁷
<i>stf1-1</i>	yes	partial†

Cells were exposed to 50 Gy γ -rays or 10 mM hydroxyurea as described in legends to Figs 1 and 2 and in ref. 7. Original data are available on request. ND, not determined.

* Human T-cell protein tyrosine phosphatase integrated into *cdc25-22 leu1-32* (ref. 12).

† Cells delay and elongate but accumulate with septa and die more rapidly than wild type.

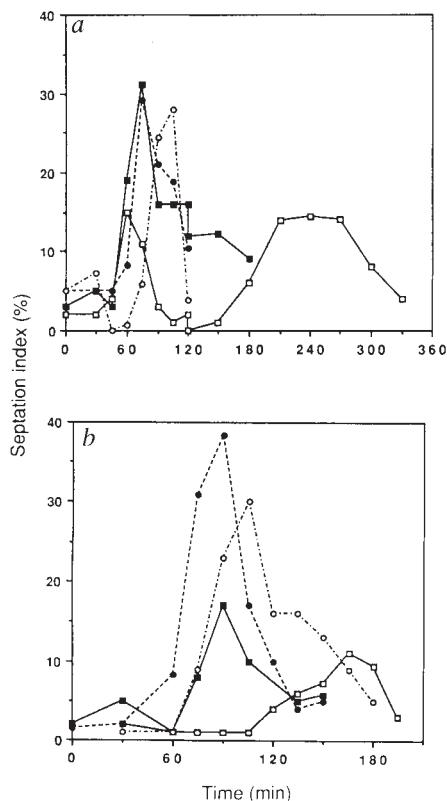


FIG. 3 *a*, Septation indices plotted against time after irradiation (100 Gy) for synchronous cultures of *wee1-50* cells. Solid lines, cells maintained at room temperature (■, non-irradiated; □, irradiated); broken lines, cells maintained at 36 °C (●, non-irradiated; ○, irradiated). Cells were synchronized by centrifugal elutriation. Non-irradiated cells at room temperature showed a peak of septation at 75 min after irradiation. Irradiated cells at room temperature were mainly delayed in septation (peak 240 min after irradiation). A small peak of septation at 60 min after irradiation probably represents cells irradiated when already past the γ -ray transition point (the last point before mitosis at which a cell may be delayed by irradiation). Cells maintained at 36 °C showed peaks of septation at 75 min and ~100 min for non-irradiated and irradiated cells respectively; thus little delay was induced when cells were maintained at the restrictive temperature for *wee1* function. An increase in cell number followed each peak of septation (not shown). *b*, Septation indices plotted against time after irradiation (100 Gy) for synchronous populations of the strains *cdc10-129h+* (solid lines; ■, unirradiated; □, irradiated) and *wee1-50 cdc10-129h-* (broken lines; ●, unirradiated; ○, irradiated). In *wee1+* cells, septum formation was delayed by ~75 min (peak-to-peak interval). In *wee1-* cells, septum formation was delayed by no more than 15 min, consistent with the small decrease in septation observed in irradiated asynchronous populations (Fig. 1c). Nuclear division, quantified by scoring the percentage of binucleate (aseptate) cells (ethanol-fixed and stained with 4'6-diamidino-2-phenylindole to reveal the nucleus) occurred coordinately with septum formation in both cell strains (not shown).

METHODS. For synchronization by elutriation, the elutriator chamber was loaded at room temperature with exponentially growing cells, then the flow rate of the medium entering the chamber was increased to retrieve the smallest single cells in the population. These cells were irradiated (100 Gy) or mock-irradiated, and samples of each shifted to 36 °C. Septation indices were monitored as in Fig. 1. To induce synchrony in *cdc10* strains, cells were maintained at 36 °C for 3 h, causing them to accumulate at the *cdc10* execution point in late G1²¹. Cells were released from this block by shifting to 25 °C; after 1 h they were irradiated or mock-irradiated. Both cell lines were then returned to 36 °C.

double mutant *wee1-50 cdc17-K42* lost viability more rapidly at 36 °C than *cdc17-K42*. The viability curves approximated to simple exponentials and had slope constants of -0.41 (correlation coefficient = 0.98) for *cdc17-K42* (36 °C) and -0.58 (correlation coefficient = 0.99) for *wee1-50 cdc17-K42* (36 °C). Similar observations were made² when comparing the responses of the budding yeast DNA ligase mutant *cdc9* with the response of the double mutant, *rad9 cdc9*.

The strains used for radiation-induced delay were also tested for their response to hydroxyurea. As previously reported⁷, the *cdc2-3w* mutation and *cdc25* overexpression both relieved the dependency of mitosis on the completion of DNA synthesis but inactivation of *wee1* did not. In addition, we found that both *mik1::ura4* and *stf1-1* mutations rendered cells unable fully to regulate progression in response to DNA synthesis inhibition. Both mutants continued septation in the presence of hydroxyurea and were abnormally likely to be killed by this

agent. Similar sensitivity was recently demonstrated in the mutant *pim1* (ref. 30), a fission yeast homologue of the human regulator of chromosome condensation RCC1 (ref. 31).

These observations clearly distinguish between the checkpoint cell cycle controls responsive to DNA synthesis and to DNA damage. The induction of a cell-cycle delay by inhibition of DNA synthesis requires that *cdc2* remain subject to regulation by *cdc25* (ref. 7), but does not require the function of the negative cell cycle regulator *wee1*. Judged by mutant responses to hydroxyurea, delay induction may also require both *mik1* and *stf1* functions. The induction of delay by irradiation, presumably due to DNA damage, requires that *wee1* be functional but is not affected by dissociating *cdc2* from regulation by *cdc25*. Radiation-induced cell-cycle delay is therefore most probably imposed by activation of *wee1*, presumably by progression controls responsive to DNA damage, like *RAD9* (ref. 4). □

Received 18 November 1991; accepted 13 January 1992.

1. Tobey, R. A. *Nature* **254**, 245-247 (1975).
2. Hartwell, L. H. & Weinert, T. A. *Science* **246**, 629-634 (1989).
3. Enoch, T. & Nurse, P. *Cell* **65**, 921-923 (1991).
4. Weinert, T. A. & Hartwell, L. H. *Science* **241**, 317-322 (1988).
5. Weinert, T. A. & Hartwell, L. H. *Cell* **12** (suppl.), 145-148 (1989).
6. Weinert, T. A. & Hartwell, L. H. *Molec. cell. Biol.* **10**, 6554-6564 (1990).
7. Enoch, T. & Nurse, P. *Cell* **60**, 665-673 (1990).
8. Nurse, P. & Thuriaux, P. *Genetics* **96**, 627-637 (1980).
9. Hindley, J. & Phear, G. A. *Gene* **31**, 129-134 (1984).
10. Russell, P. & Nurse, P. *Cell* **45**, 145-153 (1986).
11. Gould, K. G. & Nurse, P. *Nature* **342**, 39-45 (1989).
12. Gould, K. L., Moreno, S., Tonks, N. K. & Nurse, P. *Science* **250**, 1573-1576 (1990).
13. Strausfeld, U. *et al. Nature* **351**, 242-245 (1991).
14. Moreno, S. & Nurse, P. *Nature* **351**, 194 (1991).
15. Kumagai, A. & Dunphy, W. G. *Cell* **64**, 903-914 (1991).
16. Russell, P. & Nurse, P. *Cell* **49**, 559-567 (1987).
17. Featherstone, C. & Russell, P. *Nature* **349**, 808-811 (1991).
18. Russell, P. & Nurse, P. *Cell* **49**, 569-576 (1987).
19. Feilott, H., Nurse, P. & Young, P. G. *Genetics* **127**, 309-318 (1991).
20. Lundgren, K. *et al. Cell* **64**, 1111-1122 (1991).
21. Fleig, U. N. & Nurse, P. *Molec. gen. Genet.* **226**, 432-440 (1991).
22. Fantes, P. A. in *Molecular Biology of the Fission Yeast* (eds Nasim, A., Young, P. G. & Johnson, B. F.) 127-204 (Academic, San Diego, 1989).

23. Krakoff, I. H., Brown, N. C. & Reichard, P. *Cancer Res.* **28**, 1559-1565 (1968).
24. Weibezahn, K. F., Lohrer, H. & Herrlich, P. *Mutat. Res.* **145**, 177-183 (1985).
25. Leeper, D. B., Schneiderman, M. H. & Dewey, W. C. *Radiat. Res.* **53**, 326-337 (1973).
26. Doida, Y. & Okada, S. *Radiat. Res.* **38**, 513-529 (1969).
27. Mitchison, J. M. in *Methods in Cell Physiology* Vol. 4 (ed. Prescott, D. M.) 131-165 (Academic Press, 1970).
28. Hudson, J. D., Feilott, H. & Young, P. G. *Genetics* **126**, 309-315 (1990).
29. Nasmyth, K. A. *Cell* **12**, 1109-1120 (1977).
30. Matsumoto, T. & Beach, D. *Cell* **66**, 347-360 (1991).
31. Uchida, S. *et al. Molec. cell. Biol.* **10**, 577-584 (1990).
32. Gut, H., Heslot, H., Leupold, U. & Loprieno, N. in *Handbook of Genetics* Vol. 1. (ed. King, R. C.) 395-446 (Plenum, New York, 1974).

ACKNOWLEDGEMENTS. The wild-type cell line was obtained from the NRS of Canada, courtesy of R. Latta. Mutant strains and use of a centrifugal elutriator were provided by P. Russell, Scripps Institute; other mutants were donated by P. Fantes, Edinburgh University, D. Beach, Cold Spring Harbor Laboratories, and K. Gould, Vanderbilt University. Technical assistance was provided by L. A. Kort, G. Schell and C. Lingner. This work was supported by grants from the National Cancer Institute, DHHA (R.R.) and from NSERC and NCIC (P.G.Y.).

SUPPLEMENTARY INFORMATION. Original data for Table 1 are available from the London office of *Nature*.

Role of poly(ADP-ribose) formation in DNA repair

Masahiko S. Satoh & Tomas Lindahl

Imperial Cancer Research Fund, Clare Hall Laboratories, South Mimms, Herts EN6 3LD, UK

THE abundant nuclear enzyme poly(ADP-ribose) polymerase catalyses the synthesis of poly(ADP-ribose) from nicotinamide adenine dinucleotide (NAD⁺)¹⁻⁵. This protein has an N-terminal DNA-binding domain containing two zinc-fingers, which is linked to the C-terminal NAD⁺-binding domain by a short region containing several glutamic acid residues that are sites of auto-poly(ADP-ribosylation)⁶⁻⁸. The intracellular production of poly(ADP-ribose) is induced by agents that generate strand interruptions in DNA⁷. The branched homopolymer chains may attain a size of 200–300 residues⁹ but are rapidly degraded after synthesis. The function of poly(ADP-ribose) synthesis is not clear, although it seems to be required for DNA repair^{10,11}. Here we describe a human cell-free system that enables the role of poly(ADP-ribose) synthesis in DNA repair to be characterized. The results indicate that unmodified polymerase molecules bind tightly to DNA strand breaks; auto-poly(ADP-ribosylation) of the protein then effects its release and allows access to lesions for DNA repair enzymes.

γ -Irradiated plasmid DNA containing one single-strand break per molecule was incubated with soluble extracts of cultured human cells for NAD⁺-promoted rejoining of radiation-induced strand breaks. A related approach of adding damaged plasmids to human cell-free extracts was used previously for investigating DNA repair of pyrimidine dimers¹². Only a small minority of the radiation-induced strand breaks could be rejoined directly by DNA ligase owing to the presence of oxidized derivatives of deoxyribose and 3'-phosphate groups at DNA termini. Thus, repair of strand interruptions requires at least three different types of enzyme: trimming functions such as exonucleases or phosphatases to remove various terminal groups blocking DNA synthesis and ligation, a DNA polymerase for gap-filling, and a DNA ligase. The irradiated, open circular DNA was added to extracts prepared under mild conditions¹³ from human cultured lymphoblastoid cells or HeLa cells supplemented with Mg²⁺, the four deoxynucleoside triphosphates, ATP, and an

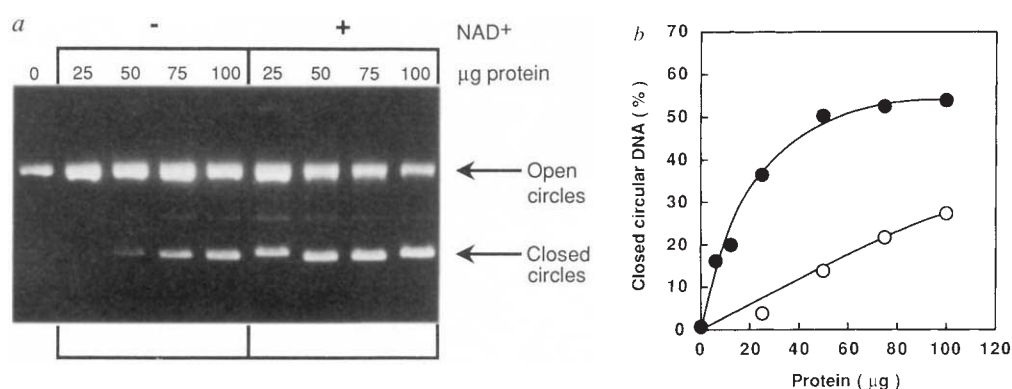
ATP-regenerating system. Conversion to covalently closed circular DNA, which reflects DNA repair, was measured by agarose gel electrophoresis in the presence of ethidium bromide. A small amount of DNA rejoining was observed in this system, especially at high protein concentrations. However, the reaction was strongly stimulated by the addition of 0.1–2 mM NAD⁺ (Fig. 1), an effect that could not be achieved by NADP⁺, NADH, or NADPH. A maximum of ~60% of the open circular DNA molecules could be converted to a closed form, reflecting the heterogeneity of γ -ray-induced DNA lesions.

Mammalian DNA ligases do not use NAD⁺ as a co-factor unlike the DNA ligase from *Escherichia coli*, neither is it known to be required by other enzymes participating in mammalian DNA repair or replication. The promotion of rejoining of γ -irradiated DNA by NAD⁺ might therefore be due to its role as a precursor in poly(ADP-ribose) synthesis. Repair of ionizing radiation damage to DNA in intact cells is inhibited by 3-aminobenzamide^{14,15}. The repair of DNA associated with NAD⁺ was also suppressed by 3-aminobenzamide in our *in vitro* system (Fig. 2a). We also found that cell extracts in which γ -irradiated open circular DNA was converted to a covalently closed form contained active poly(ADP-ribose) polymerase (relative molecular mass, 113,000; 1.8 nmol NAD⁺ incorporated into acid-insoluble material per min mg protein): poly(ADP-ribose) was synthesized in association with the repair reaction but was inhibited by 3-aminobenzamide (Fig. 2a).

One function proposed for poly(ADP-ribose) is the modification of histones to relieve chromatin condensation^{3,4}. But our cell extracts were essentially histone-free as judged by SDS-polyacrylamide gel electrophoresis, and the damaged DNA assayed was purified plasmid DNA. It is unlikely that this irradiated DNA would assemble into nucleosomes under our conditions, but we depleted cell extracts of any remaining nucleosome core histones by chromatography on phosphocellulose in 0.6 M NaCl (ref. 16). These histone-depleted extracts fully retained their ability to repair γ -irradiated DNA in an NAD⁺-stimulated reaction (data not shown). No radiolabelled histones were detected by SDS-polyacrylamide gel electrophoresis and autoradiography in cell extracts incubated with γ -irradiated DNA and [³²P]NAD.

The principal endogenous acceptor of poly(ADP-ribose) in human cell extracts was the poly(ADP-ribose) polymerase undergoing automodification (Fig. 3). Enzyme-bound polymer

FIG. 1 NAD⁺ enhancement of repair of strand breaks in γ -irradiated DNA by a human cell extract. The 3-kb pBluescript II KS⁺ plasmid (Stratagene) was propagated in *Escherichia coli* JM109. Covalently closed circular plasmid DNA (3.4 mg ml⁻¹) in 10 mM Tris-HCl, pH 8.0, 1 mM EDTA, was exposed to 50 Gy of γ -rays at 0 °C from a ⁶⁰Co source (16 Gy min⁻¹) to convert ~20% of the DNA to an open circular form. The plasmid molecules containing strand breaks were purified by two cycles of ethidium bromide/CsCl density gradient centrifugation, followed by isopropanol extraction and dialysis, and were incubated with gently prepared human cell extracts^{12,13}. Experiments were done with extracts from the normal human lymphoblastoid cell line GMO6315A (from NIGMS Human Mutant Cell Repository, NJ) propagated in RPMI 1640 medium supplemented with 15% fetal bovine serum and antibiotics. Reaction mixtures (50 μ l) contained 50 mM HEPES-KOH, pH 7.8, 5 mM MgCl₂, 70 mM KCl, 0.5 mM dithiothreitol, 0.4 mM EDTA, 20 μ M each of dATP, dCTP, dGTP, and TTP, 2 mM ATP, 40 mM phosphocreatine, 2.5 μ g creatine phosphokinase (type I, Sigma), either 0 or 2 mM NAD⁺, 0.3 μ g γ -irradiated open circular DNA, and cell extract (0–100 μ g). Incubations were for 3 h at 30 °C. Reactions were terminated by addition of SDS to 0.6% and EDTA to 20 mM. After digestion with proteinase K (240 μ g ml⁻¹)



for 30 min at 37 °C and extraction with phenol/chloroform, the amount of DNA converted to a covalently closed circular form was determined by ethidium bromide/agarose gel electrophoresis. DNA bands were visualized by ultraviolet light (a) and photographic negatives analysed by microdensitometry; signals for covalently closed circular DNA were corrected by a factor of 1.45 to compensate for the reduced binding of ethidium bromide (b). The amount of closed circular DNA was expressed as a per cent of total plasmid DNA. The faint band between the closed and open circular DNA bands comigrated with full-length linear plasmid DNA. Filled symbols, reaction mixtures with NAD⁺; open symbols, reaction mixtures without NAD⁺.

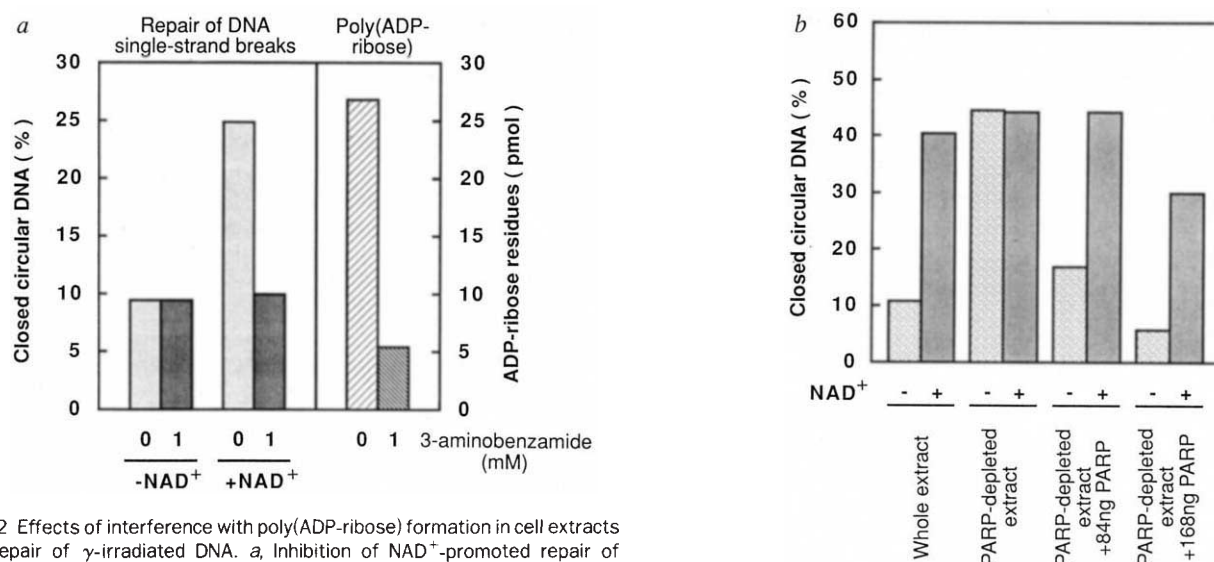


FIG. 2 Effects of interference with poly(ADP-ribose) formation in cell extracts on repair of γ -irradiated DNA. *a*, Inhibition of NAD⁺-promoted repair of γ -irradiated DNA and poly(ADP-ribose) synthesis by 3-aminobenzamide. Reaction mixtures contained 25 μ g cell extract protein and were as described in Fig. 1 legend, except that the NAD⁺ concentration was 0.25 mM and the incubation time 30 min. Reactions were supplemented with 1 mM 3-aminobenzamide (Sigma). For analysis of poly(ADP-ribose) formation, reaction mixtures (containing 0.25 mM NAD⁺) were supplemented with 2 μ Ci [³²P]NAD (Amersham). In this case, reactions were terminated by addition of 0.75 ml cold 20% trichloroacetic acid. Acid-insoluble polymeric material was collected on glass filters, washed twice with 5% trichloroacetic acid and ethanol, and radioactivity determined by liquid scintillation counting. Standard errors between three different experiments were below 20%.

was rapidly synthesized by extracts containing NAD⁺ and γ -irradiated DNA, the amount of poly(ADP-ribose) reached a maximum after 10–30 min incubation and was followed by degradation. The automodification of the polymerase decreases the binding affinity of the protein for DNA^{17–19}, presumably by electrostatic repulsion between the two negatively charged polymers. Thus, the rapid poly(ADP-ribose) synthesis occurring in connection with DNA repair would be expected to lead to release of the polymerase from the damaged DNA. We depleted human cell extracts of this enzyme by chromatography on double-stranded DNA-cellulose in 0.4 M NaCl (ref. 20). The enzyme-depleted extracts efficiently catalysed repair of γ -irradiated DNA in the absence as well as in the presence of NAD⁺ (Fig. 2*b*), showing that removal of the polymerase facilitated the rejoining of strand breaks in γ -irradiated DNA. In reconstitution experiments, supplementing depleted extracts with purified polymerase reestablished the dependence of effective DNA repair on NAD⁺ (Fig. 2*b*). Addition of purified DNA ligases or polymerases to reaction mixtures indicated that the polymerization or damage-excision steps, rather than DNA ligation, were the main targets of inhibition by poly(ADP-ribose) polymerase in the absence of NAD⁺ (data not shown).

A model based on our results proposes that poly(ADP-ribose) participates in the repair of γ -irradiated DNA (Fig. 4), as has already been suggested¹⁸. The poly(ADP-ribose) polymerase binds strongly to strand breaks in DNA and interferes with the access of repair enzymes to the damaged site. In the presence of inhibitors of poly(ADP-ribose) synthesis, such as 3-aminobenzamide, this situation may persist. Otherwise an intrinsic release mechanism of the enzyme triggers rapid automodification so that the automodified poly(ADP-ribose) polymerase no longer binds to DNA breaks, thereby making the damaged site available for repair.

It has been noted before that automodification of poly(ADP-ribose) polymerase decreases its affinity for DNA and a shuttle mechanism was proposed for modulating DNA-protein interactions, in which the polymerase can be moved off and on DNA by automodification and polymer degradation by a glycohy-

drolase, respectively¹⁹. In a 'mono-polymerase' cell-free system for DNA replication²¹, elongation of DNA replication intermediates was inhibited by a protein, identified as the poly(ADP-ribose) polymerase. Poly(ADP-ribosylation) of DNA polymerases and DNA ligases *in vitro* only leads to inhibition of their activities²².

Our results can explain why poly(ADP-ribose) is formed during DNA repair, but not the main function of the poly(ADP-ribose) polymerase, which does not seem to be obligatory for

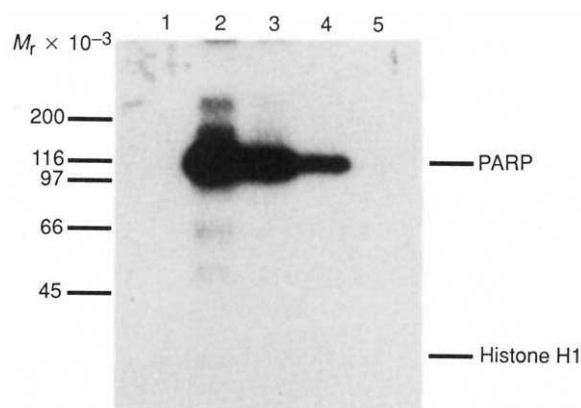


FIG. 3 Automodification of poly(ADP-ribose) polymerase (PARP) during repair of γ -irradiated DNA in a human cell extract. Complete reaction mixtures were described for Fig. 1 and contained 25 μ g protein extract but only 32 nM [³²P]NAD (2 μ Ci). Reactions were incubated for various times and stopped by addition of 55 μ l ice-cold acetone. Protein precipitates were collected by centrifugation and analysed by SDS-polyacrylamide gel electrophoresis and autoradiography. Molecular weight (M_r) markers (Bio-Rad) are shown on the left. Samples were incubated for 0 min (lane 1), 10 min (lane 2), 60 min (lane 3) and 120 min (lane 4). In lane 5, a 10-min incubation of a reaction mixture containing non-irradiated covalently closed circular plasmid DNA is shown. Modifications at high NAD⁺ concentrations (250 μ M) yielded qualitatively similar results, except that longer poly(ADP-ribose) chains were formed, which delayed protein migration.

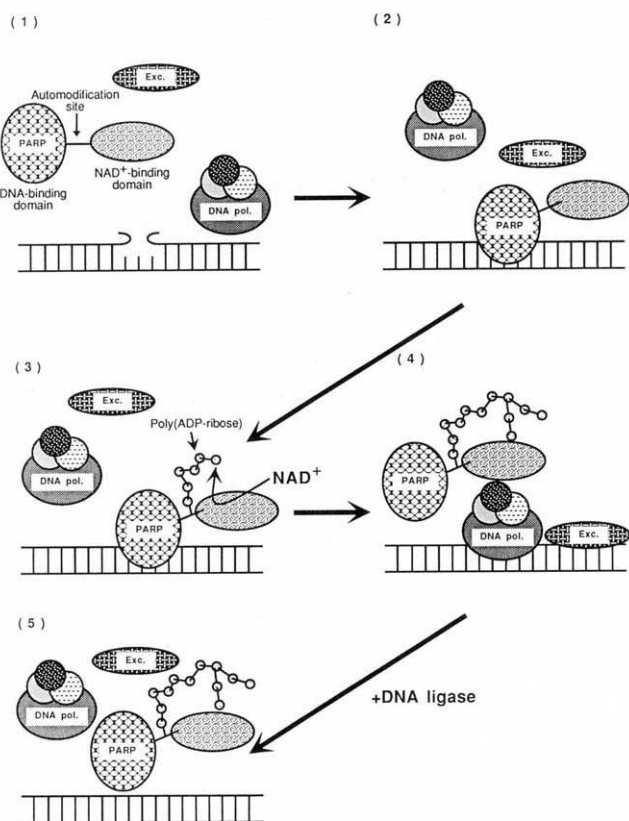


FIG. 4 Model for the involvement of poly(ADP-ribose) formation in DNA repair. After introduction of a DNA strand interruption with frayed termini by γ -irradiation, PARP, excision enzymes (Exc.) and DNA polymerases (DNA pol.) compete for binding to the damaged site (1); attachment of PARP interferes with access to the strand break for repair enzymes (2); poly(ADP-ribose) synthesis is activated by binding of PARP to damaged DNA (3); the automodified PARP has reduced affinity for DNA and is released, allowing availability of the lesion to DNA-repair enzymes (4); excision-repair seals the DNA strand break (5). Unmodified PARP is regenerated by degradation of poly(ADP-ribose) with a specific glycohydrolase.

DNA repair (Fig. 2b). One practical application could be to enhance the cytotoxic effects of DNA-damaging drugs and radiation by inhibiting poly(ADP-ribose) formation with nicotinamide analogues, given the paucity of specific inhibitors of DNA repair. The main function of the polymerase may not even be associated with DNA repair—the enzyme is abundant in chromatin and so could have a structural role in cell nuclei. Poly(ADP-ribose) is found in all higher and many lower eukaryotes but not in yeast or bacteria, although yeast nucleosomes are like those in higher organisms. On the other hand, the dinoflagellate *Cryptocodinium cohnii* lacks histones and has no nucleosome organization, but still has a polymerase for poly(ADP-ribose) synthesis and a poly(ADP-ribose) glycohydrolase²³. These observations, together with our data, make it unlikely that histone modification is the main task of poly(ADP-ribose) polymerase, although chromatin decondensation may be achieved by modification of histone H1 in heterochromatin²⁴. Poly(ADP-ribose) polymerase could help DNA strand breaks introduced temporarily for a purpose¹¹, for example in DNA associating with the nuclear scaffold, at origins of replication, or during cellular differentiation. Such a 'nick-protection' mechanism might prevent recombination leading to chromosome aberration⁵ and spurious initiation of transcription at DNA strand breaks²⁵. Alternatively, the binding of the poly(ADP-ribose) polymerase to strand breaks in DNA may constitute a signal that switches off DNA synthesis in γ -irradiated mammalian cells to ensure that lesions are not replicated before repair. □

Received 21 November 1991; accepted 15 January 1992.

- Chambon P., Weill J. D., Doly J., Strosser M. T. & Mandel P. *Biochem. biophys. Res. Commun.* **25**, 638–643 (1966).
- Sugimura T. *Progr. Nucleic Acid Res. molec. Biol.* **13**, 127–151 (1973).
- Ueda K. & Hayaishi O. A. *Rev. Biochem.* **54**, 73–100 (1985).
- Althaus F. R. & Richter C. *ADP-Ribosylation of Proteins, Enzymology and Biological Significance* 1–125 (Springer, Berlin, 1987).
- Cleaver J. E. & Morgan W. F. *Mut. Res.* **257**, 1–18 (1991).
- Kameshita I., Matsuda Z., Taniguchi T. & Shizuta Y. *J. biol. Chem.* **259**, 4770–4776 (1984).
- Gradwohl G. et al. *Proc. natn. Acad. Sci. U.S.A.* **87**, 2990–2994 (1990).
- Desmarais Y., Menard L., Laguerre J. & Poirier G. G. *Biochim. biophys. Acta* **1078**, 179–186 (1991).
- Miwa M. & Sugimura T. *Meth. Enzym.* **106**, 441–450 (1984).
- Durkacz B. W., Omidji O., Gray D. A. & Shall S. *Nature* **283**, 593–596 (1980).
- Shall S. *Adv. Radiat. Biol.* **11**, 1–69 (1984).
- Wood R. D., Robins P. & Lindahl T. *Cell* **53**, 97–106 (1988).
- Manley J. L., Fire A., Samuels M. & Sharp P. A. *Meth. Enzym.* **101**, 568–582 (1983).
- Nduka N., Skidmore C. J. & Shall S. *Eur. J. Biochem.* **105**, 525–530 (1980).
- Ben-Hur E., Utsumi H. & Elkind M. M. *Radiat. Res.* **97**, 546–555 (1984).
- Smith S. & Stillman B. *EMBO J.* **10**, 971–980 (1991).
- Yoshihara K. et al. *J. biol. Chem.* **256**, 3471–3478 (1981).
- Ferro A. M. & Olivera B. M. *J. biol. Chem.* **257**, 7808–7813 (1982).
- Zahradka P. & Ebisuzaki K. *Eur. J. Biochem.* **127**, 579–585 (1982).
- Zahradka P. & Ebisuzaki K. *Eur. J. Biochem.* **142**, 503–509 (1984).
- Eki T. & Hurwitz J. *J. biol. Chem.* **266**, 3087–3100 (1991).
- Yoshihara K. et al. *Biochem. biophys. Res. Commun.* **128**, 61–67 (1985).
- Werner E. et al. *Eur. J. Biochem.* **139**, 81–86 (1984).
- Poirier G. G., de Murcia G., Jongstra-Bilen J., Niedergang C. & Mandel P. *Proc. natn. Acad. Sci. U.S.A.* **79**, 3423–3427 (1982).
- Slattery E., Dignam J. D., Matsui T. & Roeder R. G. *J. biol. Chem.* **258**, 5955–5959 (1983).

ACKNOWLEDGEMENTS. We thank G. Poirier, S. Sato, S. Shall and R. Wood for discussion, and L. Jones for help with γ -irradiation experiments. M.S.S. was a recipient of a Research Training Fellowship awarded by the International Agency for Research on Cancer. This work was supported by the Imperial Cancer Research Fund.

Suppression of a myosin defect by a kinesin-related gene

Sue H. Lillie & Susan S. Brown

Department of Anatomy and Cell Biology,
University of Michigan Medical School, Ann Arbor, Michigan 48109, USA

MOTOR proteins in cells include myosin¹, which is actin-based, and kinesin, dynein and dynamin², which are microtubule-based. Several proteins have recently been identified that have amino-acid sequences with similarity to the motor domains of either myosin³ or kinesin^{4–9}, but are otherwise dissimilar. This has led to the suggestion that these may all be motor proteins, but that they are specialized for moving different cargos. Genetic analysis can address the question of the different functions of these new proteins. Studies of a temperature-sensitive mutation (*myo2-66*) in a gene of the myosin superfamily (*MYO2*) have implicated the Myo2 protein (Myo2p) in the process of polarized secretion in yeast (*Saccharomyces cerevisiae*)¹⁰. To understand more about the role of Myo2p, we have looked for 'multicopy suppressors' (heterologous genes that, when overexpressed, can correct the temperature sensitivity of the *myo2-66* mutant). Here we report the identification of such a suppressor (*SMY1*) that (surprisingly) encodes a predicted polypeptide sharing sequence similarity with the motor portion of proteins in the kinesin superfamily.

SMY1 (suppressor of myosin) was isolated by introducing a yeast genomic library, in a high-copy-number plasmid, into a *myo2-66* mutant strain and looking for yeast transformants that could grow at restrictive temperature (Fig. 1). Phenotypic analysis revealed that although *SMY1* is a relatively good multicopy suppressor, the suppression is not complete (Fig. 2). The lethality of the *myo2-66* mutation is suppressed at 30–33 °C but not at 35 °C. At 31 °C, the doubling time of the *myo2-66* strain suppressed by plasmid-borne *SMY1* is three to four times longer than that of the strain carrying plasmid-borne *MYO2*. The strain with vector alone undergoes less than one doubling before arresting as large unbudded cells at ≥ 30 °C, as expected¹⁰.

The region of DNA sufficient for suppression was localized to a 2.3-kilobase (kb) fragment (Fig. 1). Sequencing of this region (Fig. 3) revealed a single long open reading frame. The

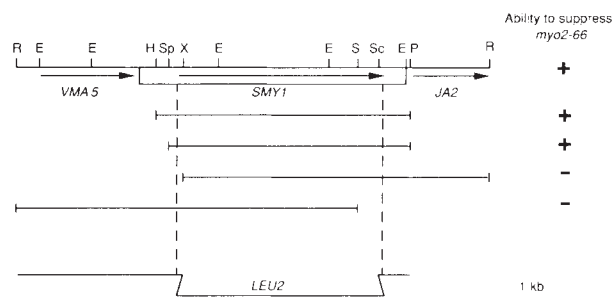


FIG. 1 Map of the *SMY1* locus. Top line, a subcloned segment of the *SMY1*-containing region obtained from the genomic library; other subcloned segments are aligned beneath it, and their ability to suppress the *myo2-66* mutation is indicated. Open box, the region whose sequence is shown in Fig. 3. Arrows, open reading frames identified by sequencing. *VMA5* (which encodes a 40K subunit of the vacuolar ATPase; Tom Stevens, personal communication) and *JAZ2* (ref. 23 and J. Arenas, personal communication) lie just 400 bp upstream, and 280 bp downstream of *SMY1*, respectively (the positions of the distal ends of these flanking genes shown in the figure are not exact). Bottom line, the construct used for *SMY1* deletion; the 1.9-kb *XbaI*-*ScaI* fragment of *SMY1* was replaced with a 2-kb fragment containing the *LEU2* gene. Vertical dashed lines, delineate the open reading frame. All restriction enzyme sites are shown for: E, *EcoRI*; H, *HpaI*; P, *PstI*; R, *EcoRV*; S, *SstI*; Sc, *ScaI*; Sp, *SpeI*; X, *XbaI*.

METHODS. The *SMY1*-containing plasmid was obtained by transforming a *myo2-66 ura3-52* strain with the yeast library in YEp24 (a high-copy-number *URA3*-containing plasmid²⁴). Transformation plates, selective for the *URA3* auxotrophic marker, were incubated at permissive temperature (~20 °C) for 2 days and then shifted to restrictive temperature (31 °C). The plasmid was recovered from a yeast transformant into *Escherichia coli*; restriction fragments were subcloned into the high-copy-number plasmid YEp352²⁵ and tested for ability to suppress a *myo2-66* mutant strain. This library screen only yielded two other temperature-sensitive⁺ (*ts*⁺) transformants out of the 16,000 *Ura*⁺ transformants screened. Although this was not an exhaustive search, it does indicate that there are not a large number of genes capable of multicopy suppression of *myo2-66* (based both on a calculated expected frequency, and on observed frequencies in other multicopy suppressor screens with this library; S.H.L. *et al.*, unpublished results). *MYO2* was not recovered in this screen, presumably because its overexpression is deleterious (S.H.L. and S.S.B., unpublished results).

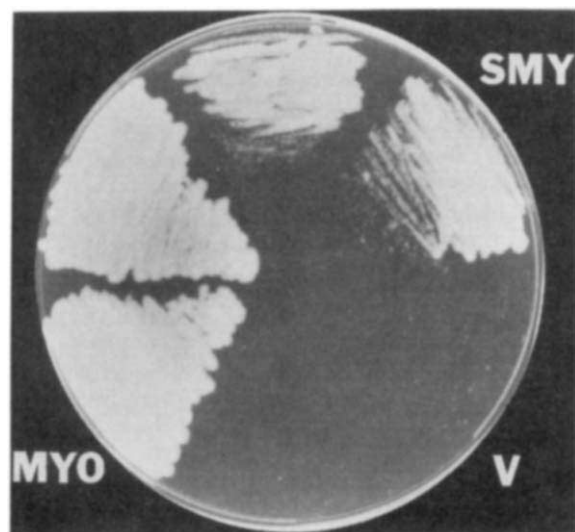


FIG. 2 *SMY1* suppression of the *myo2-66* mutation. Six independent transformants of a *myo2-66* mutant were spread on a plate and incubated at 31 °C for 36 h. Two contained the *SMY1* gene on a high-copy-number plasmid (*SMY*), two the high-copy-number vector alone (*V*), and two the *MYO2* gene on a low-copy-number plasmid (*MYO*). *MYO2* was introduced on a low-copy-number plasmid because the cells seem not to grow well with *MYO2* on a high-copy-number plasmid.

putative start codon is preceded by several stop codons in all frames within the region extending 150 base pairs (bp) upstream. The putative stop codon is followed by several in-frame stop codons. There is no recognition signal for splicing within or near the open reading frame, suggesting that this gene (like most yeast genes) does not contain introns. These considerations, together with the facts that the region sufficient for suppression starts 83 bp upstream of the putative start codon and ends 264 bp downstream of the putative stop codon, led us to conclude that the predicted protein product shown in Fig. 3 is the Smy1 protein (Smy1p). This protein has 656 amino-acid residues and a calculated relative molecular mass of 73,884.

A search of the Genbank and EMBL databases revealed that the N-terminal portion of Smy1p shows significant sequence similarity to the motor domains of the kinesin superfamily, although it is the most divergent (Fig. 4). This superfamily includes kinesins², which display sequence similarity throughout their entire lengths¹¹⁻¹³, and several kinesin-related proteins that show sequence similarity only in the motor domains⁴⁻⁹. The motor domains have a number of highly conserved short stretches, some but not all of which are found in Smy1p (Fig. 4). For example, the putative ATP-binding region, indicated by the overline in Fig. 4, diverges from that typical of other kinesin-related proteins although all fit the consensus sequence of GX₄GKT¹⁴. In fact, Smy1p does not resemble kinesin-related proteins more than it does myosins^{1,3} in the ATP-binding region. It would be interesting to know how much similarity Smy1p shows to the microtubule binding region of kinesin, but this region has not been sufficiently well localized¹³ to permit a

```

-354 TTTT TTTT TTTAG AATTTT CAAC TTTT TTTT ATTC ATAG AATA AACT GTGATT GTGAAT GTTGAATAGTAATG
-279 GAACATAATCGGATAAAATATGCTTTTATGACATTTTTCATTTTTCCTGTTGAGTTTCCATCTCTT
-204 CCTTATTTATGTTTAAACGGAAGAGGTGGAAGTTTAAATAATGGAATTAATAAAAAAATAAATAA
-129 GAAAAATAAAATAGAAAAATGGCAGCGAAATTCACAGGATTTAGCATCTAGGGGGTTTATTAAGCATACA
-54 GGTCGACTTTAAACGAGAAAAATATCATCTACGGGTAGTGGCAACTTGTTTAAATGCACTGCAATGATATTTTC
1 M H W N I I S
21 GAAAGAGCAAGTAGCTCTTCTGTATCGTTACCAACTCTAGACAGCAGTGAACCTGTGACATCGAAGTTATCT
8 K E Q S S S S S V S L P T L D S S E P C H I E V I L
96 CAGGCGGATACCGAAAAAGGATTACAAAAAATGAGTGCACCTTCAAAATAGACCATATGAAAAATCTGTGCT
33 R A I P E K G L Q N N E S T F K I D P Y E N T V L
171 ATTTCGCAACAAATCCGTTACATGAGACACCAAGGAGACCGCATTCACATTTTCGATAAAGTATTGCA
58 F R T N T N P L H E T T K E T H S T F Q F D K V F D
245 TGCTACGCCCATCAAGAAGTGTTCAGAAATTTCTGGTGCATCAATAATAAATGATGTTTGAATGGTTATAA
83 A N A T Q E D V Q K F L V H N P I N D V L N G G V N
321 CGGTACTGTATAACATATGGACCAAGTTTCAGCGGAAGTCTTATCCCTTATGATGATCAAAAGCGGAAG
109 G T V I T Y G P S P S G K S L I G S K E S E G
396 AATTCACCGAATATGCAAGACCTTGTGTATACCTAGAGAAAAATGAAGAAAAAGGAGGATGTTTATG
133 I L P N I C K T L F D T L E K N E E T K G D S F S
471 CGTAAGTGTTTTGGCATTGCAAAATATATGGAAGAAACGTATGACTTATGGTACCTTTACCTGAAAGAAAC
158 V S V L A F E I Y M E K T Y D L L V P L P R K P
546 ATTAAGCTTCCCGTTCTTCAAGCAAAATGGACTAGAAATCAAGATATTTGTCGGGACATGTCGGATCGTA
183 L K L H R S S S K M D L E I K D I C P A H V G S Y
621 TGAAGACTTAAGAGCTACATTCAGGCGAGTCCAAAGCTGCGGCAATAGGATGGCATGTGGCGCAAGACAGAG
208 E D L R S Y I Q A V Q N V G N R M A C G D K T E R
696 ATCAGATCACACCTATGTTTCAATACACGTGACCAAGGAATGAAAAATGATATATTAATAAATAGTTT
233 S F S H L V F Q L H V E N N R K I N D V L N G G V N
771 TTTATCTTGGTGTATTTACCGGGGAGAGAGTTTCATAAAGACTGAAAGACTGATGATGATGATGATGAT
258 L Y L V D L H G A E K F D K R T E S T S Q D A L
846 AAAAAATTAACCAATCTATTGAGCGTTGAAAAACACTGTGCGTCTTGTCAATGAAGAGGCGTATTCAGC
283 K K L N Q S I E A L K N T V R S L S M K E R D S A
921 CTACAGCGCAAGGATCACATAGTTCTGTACCGTGAATCGCAATTAAGTGAAGTTTAAAGATTCCTCTGG
308 Y S A K G S H S S A Y R E S Q I L T E V L K D S L G
996 GGGAAATAGGAAACAAAGGTGATATTGACATGTTCTTAAGTATGTTTCAACATCCCTATCAACATGAT
333 G N R K T K V I L T C F L S N V P T T L S T L E F
1071 TGGTGACAGTATTAGACAGATCAATAACAGGTTACAGATAACACACAGGTTTAAATCGAAAAAATAGGGA
358 G D S I R Q I N N K T V D N T T G L N L K K K M A
1146 TCTATTATTTCAGGACATGAAAAATAGGATGATATTAATGTGCGCAATTAATATCTAAAGAGGCTGAATAG
383 L F I O D M K I K D N V Y A K I N D V L N G G V N
1221 CTCTTAAAAAGCCTTCAATAAATCTCTCCGAGAGCAGCAGAAAAAATGTTAGAAAATACAAAGAAAGA
408 S L L P L H N K S L P E D D E K F M L E N T K K E
1296 AAATATCAAACTAAAGCTTCAATAGACAGTATTACCAATATTGAGTAGTTCTACCAATGAAGCTCTCAACA
433 N I K L K L Q L D S I T Q L L S S S T N E D P N N
1371 TCGTATTGACGAAGAGTTTCTGAAATATTAAAGAAAGTGGCAACAGATTGCTCAGCTTGAGTTATCTTTGA
458 R I D E E V S E I L T K R C E Q I A Q L E L S F D
1446 TCGACAGATGAATCCAAATCTAAGTTCAGCAAGAGTTAGAAATACAAAAAGTCAAAAGAGGAGGCTTAGAGT
483 R Q M N S N S K L Q Q E L E Y K K S K E E A L E S
1521 TATGAAGCTTAGGCTACTAGAACAAATTCAGCTTCAGGAAGAGAGATTCAAGAGCTTTTAACTATCAAGCCAT
508 M N V R L L E Q I O L Q E R E I O E L L T N A I
1596 CTTAAGGGTGAACTAGAACTCACATAACTTACGTAACACGAGTGAAGAGATAAAATCTTTGGAAGTTTC
533 K G E I L K H E T K L E T R S E S I S
1671 TGTCAAGAGCTTCTCTTAATAAATCTGCAATCCATCTCTGAGAGAGGCTTAGAGCTGGTTCACAGAA
558 V K E L S L N R S A I P S P R G M S S S N
1746 TACTATGTCATATCGAGGAGGTTCTGGAATATCAAAATGAGCCTTGTGAGCAATATCTCTCAAAACCTTT
583 T M L H I E E G S E I S N S P W S A N T S K S P L
1821 AGTATGGGGCGGCAAGAGTTTCTTACAGCATATAGCAACCATGGATCAGAGGATCTTTTGTGGCAAGCC
608 V W G A R K V S S S S I A T T G S Q E S F V A R P
1896 GTTCAAGAGGAGTTAACTCCATTCGATAAAGTCACTTCGAGTACTCCAAAGAGTCCCTCTTCGGAAGCTG
633 F K K G L N L H S I K V T S S S T P K S P S S G S
1971 ACTATAATTAGTTCGATAATTGACGACCAACTGTGTAGTATCTTGTAAATATAATCATGATGATACGCTA
2046 ATAAATTTGATGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT
2121 AATGATTTGGAAAAATGAAAAATGAAAAATGAAAAATGAAAAATGAAAAATGAAAAATGAAAAATG
2196 GTAAGAA

```

FIG. 3 The nucleotide and predicted amino-acid sequences of *SMY1*. Numbers start at the first ATG of the open reading frame. DNA was sequenced by the dideoxy-chain termination method²⁶, both strands were completely sequenced. In all cases we have sequenced across cloning junctions on at least one strand. The Genbank accession number is M69021.

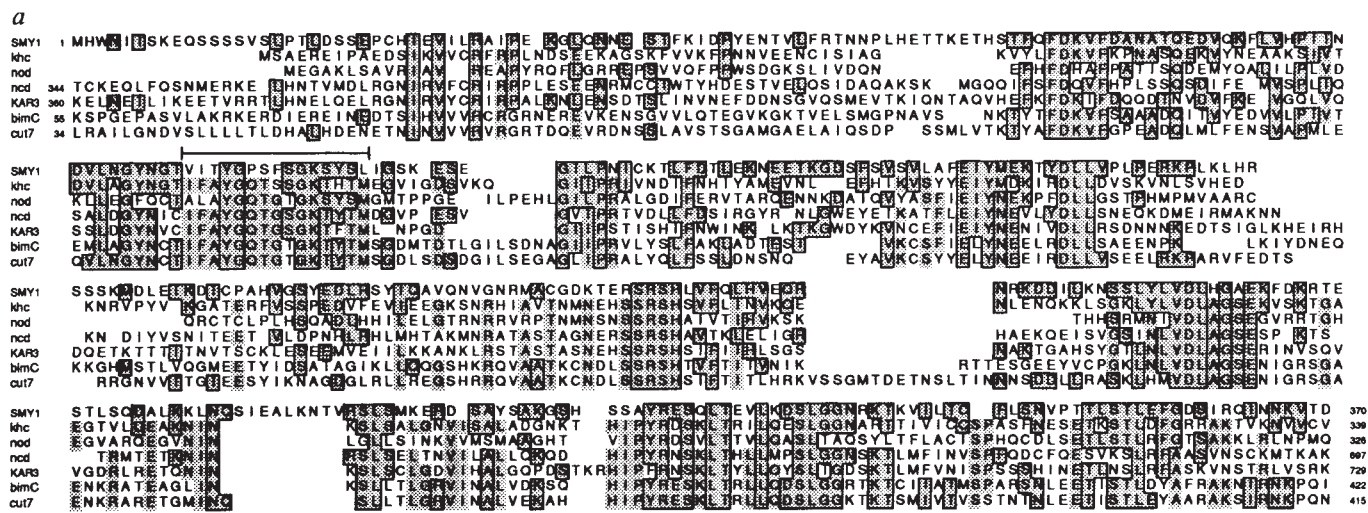


FIG. 4 Comparison of the N-terminal amino-acid sequence of Smy1p to the motor domains of kinesin and kinesin-related proteins (khc, *Drosophila* kinesin heavy chain¹³; nod, *Drosophila* nod protein⁴; nod, *Drosophila* claret nondisjunctional protein^{5,6}; KAR3, *S. cerevisiae* Kar3 protein⁷; bimC, *Aspergillus* bimC protein⁸; cut7, *Schizosaccharomyces pombe* cut7 protein⁹). The *nod* and Kar3 proteins have their motor domains at the C- instead of N-terminus. a, Residues identical to Smy1p are boxed and shaded. Residues identical in at least four proteins excluding Smy1p are shaded but not boxed. Alignments were made by a combination of BESTFIT²⁷ and visual inspection; spaces represent gaps inserted to maximize matches. Overline, putative ATP-binding region. b, Per cent identity of the sequences in a is shown in tabular form. Pairwise comparisons were performed by BESTFIT analysis, using the default parameters.

comparison. The C-terminal portion of Smy1p (the 'tail') shows no sequence similarity to kinesin or other kinesin-related proteins. We looked carefully for similarity between the Myo2p and Smy1p tails, as our results indicate that they might carry the same cargo, but none was seen.

Secondary structure predictions¹⁵ for Smy1p are similar to those for other kinesin-related proteins; amino acids 400–560 in Smy1p are predicted to be largely in an α -helical conformation, lacking turns. In conventional kinesin¹³, this predicted domain has been shown to correspond to a coiled-coil region responsible for dimerization, preceded by the motor domain, and followed by a C-terminal domain that binds light chains and probably membrane vesicles (the cargo). A computer program¹⁶ predicts that residues 272–306, 421–447 and 490–542 of Smy1p can form coiled-coils.

To learn more about the role of Smy1p, we deleted one copy of the *SMY1* gene in a diploid strain, using the construct shown in Fig. 1. This removed all but the N-terminal 16 and C-terminal 9 codons of *SMY1*. Southern blotting (data not shown) confirmed that the expected deletion had occurred in the diploid strain and the *SMY1* deletion mutant haploid (Leu⁺) progeny. Phenotypic analysis of the haploid segregants showed no detectable differences between the deletion mutants and their wild-type sisters. The deletion mutants appeared to have a normal distribution of actin and tubulin, and appeared normal in their ability to mate, sporulate and grow in various media at various temperatures. Thus, the function of Smy1p must be either redundant or nonessential under the conditions tested. If redundant, the functional equivalent is not similar enough in sequence to *SMY1* to have been detected by Southern hybridization (data not shown).

Additional genetic evidence that Smy1p may interact with or substitute for Myo2p is provided by our finding of 'synthetic lethality'¹⁷ between the *myo2-66* mutation and the *SMY1* deletion; from a cross of a *myo2-66* mutant with a *SMY1* deletion mutant no live double mutants were obtained. 36 tetrads were analysed at 20 °C (permissive temperature for the *myo2-66* mutant); these gave a frequency of 5:24:7 tetrads with 4, 3 and 1–2 live segregants, consistent with synthetic lethality. Of the 31 non-viable segregants for which the genotype could be inferred, 30 carried both mutations. The synthetic lethality indicates

that Smy1p can compensate for the defect in *myo2-66* mutants not only under the abnormal conditions of overexpression (the multicopy suppression) but also when Smy1p is presumably present at normal levels.

Thus, Smy1p, a putative microtubule-based motor, can interact with or substitute for Myo2p, a putative actin-based motor. It is possible that Myo2p functions to organize or stabilize the actin cytoskeleton. Another possibility, proposed by Johnston *et al.*¹⁰, is that Myo2p attaches to secretory vesicles and carries them along actin filaments to the bud. If a defect in vesicle transport is being corrected by *SMY1*, an obvious possibility is that Smy1p can carry the vesicles along microtubules, which are in the right location in yeast to direct vesicles to the bud. However, it has been demonstrated in yeast that microtubules are not required for delivery of secretory vesicles to the bud^{17,18}, nor do they substitute in actin mutants¹⁹. Nonetheless, microtubules could normally provide a minor pathway that can compensate for a partial defect in myosin at the nominally permissive temperature; this pathway may become more important under conditions of Smy1p overexpression, compensating for the more severe myosin defect at restrictive temperature. Such transport of vesicles along microtubules by Smy1p would represent the first clear-cut example of functional redundancy between actin and microtubules.

Another possibility is that Smy1p suppresses the myosin defect by carrying vesicles along actin filaments, although this seems unlikely, given the sequence similarity of Smy1p to kinesin. However, sequence similarity should be interpreted with caution, as shown by the recent finding that a kinesin-like protein has a directionality of movement *in vitro* opposite that of kinesin^{20,21}. Furthermore, not enough is known about actin- or microtubule-binding sequences to look for them in Smy1p; studies of truncated kinesins¹³ indicate only that the microtubule binding site(s) is somewhere in the C-terminal half of the motor domain. Thus, it will be important to test both Smy1p and Myo2p for motor activity, and to determine whether they move along actin or microtubules. As yeast actin and tubulins are similar in sequence to their mammalian counterparts and can copolymerize with them²², it is unlikely that any unusual properties of these proteins are responsible for the intriguing relationship that we have found between Smy1p and Myo2p. □

Received 4 November 1991; accepted 28 January 1992.

- Warrick, H. M. & Spudich, J. A. *Rev. Cell Biol.* **3**, 379–421 (1987).
- Vallee, R. B. & Shpetner, H. S. *A. Rev. Biochem.* **59**, 909–932 (1990).
- Pollard, T. D., Doberstein, S. K. & Zot, H. G. *A. Rev. Physiol.* **53**, 653–681 (1991).
- Zhang, P., Knowles, B. A., Goldstein, L. S. B. & Hawley, R. S. *Cell* **62**, 1053–1062 (1990).
- Endow, S. A., Henikoff, S. & Soler-Niedziela, L. *Nature* **345**, 81–83 (1990).
- McDonald, H. B. & Goldstein, L. S. B. *Cell* **61**, 991–1000 (1990).
- Meluh, P. B. & Rose, M. D. *Cell* **60**, 1029–1041 (1990).
- Enos, A. P. & Morris, N. R. *Cell* **60**, 1019–1027 (1990).
- Hagan, I. & Yanagida, M. *Nature* **347**, 563–566 (1990).
- Johnston, G. C., Prendergast, J. A. & Singer, R. A. *J. Cell Biol.* **113**, 539–551 (1991).
- Kosik, K. S., Orecchio, L. D., Schnapp, B., Inouye, H. & Neve, R. L. *J. Biol. Chem.* **265**, 3278–3283 (1990).
- Wright, B. D. *et al. J. Cell Biol.* **113**, 817–833 (1991).
- Yang, J. T., Laymon, R. A. & Goldstein, L. S. B. *Cell* **56**, 879–889 (1989).
- Walker, J. E., Saraste, M., Runswick, M. J. & Gay, N. J. *EMBO J.* **1**, 945–951 (1982).
- Garnier, J., Osguthorpe, D. J. & Robson, B. *J. molec. Biol.* **120**, 97–120 (1978).
- Lupas, A., Van Dyke, M. & Stock, J. *Science* **252**, 1162–1164 (1991).
- Huffaker, T. C., Hoyt, M. A. & Bostein, D. A. *Rev. Genet.* **21**, 259–284 (1987).
- Jacobs, C. W., Adams, A. E. M., Szaniszlo, P. J. & Pringle, J. R. *J. Cell Biol.* **107**, 1409–1426 (1988).
- Novick, P. & Bostein, D. *Cell* **40**, 405–416 (1985).
- Walker, R. A., Salmon, E. D. & Endow, S. A. *Nature* **347**, 780–782 (1990).
- McDonald, H. B., Stewart, R. J. & Goldstein, L. S. B. *Cell* **63**, 1159–1165 (1990).
- Drubin, D. *Cell. Motil. Cytoskel.* **14**, 42–49 (1989).
- Company, M., Arenas, J. & Abelson, J. *Nature* **349**, 487–493 (1991).
- Carlson, M. & Botstein, D. *Cell* **28**, 145–154 (1982).
- Hill, J. E., Myers, A. M., Koerner, T. J. & Tzagoloff, A. *Yeast* **2**, 163–167 (1986).
- Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463–5467 (1977).
- Devereux, J., Haeberli, P. & Smithies, O. *Nucleic Acids Res.* **12**, 387–395 (1984).

ACKNOWLEDGEMENTS. We thank G. Johnston and R. Singer for *myo2-66* mutant strains, *MYO2* DNA, and for communicating unpublished results, T. McKinney for sequencing assistance, and D. Gumucio, B. Haarer and J. Schiefelbein for critical reading of the manuscript. This work was supported by a grant from the American Cancer Society and in part by an NIH grant to the Kughn Clinical Research Center at the University of Michigan Hospital.

Diode-like behaviour of a mitochondrial electron-transport enzyme

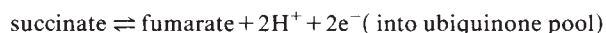
Artur Sucheta*, Brian A. C. Ackrell†, Bruce Cochran† & Fraser A. Armstrong*

*Department of Chemistry, University of California, Irvine, California 92717, USA

†Veterans Administration Hospital and Department of Biochemistry and Biophysics, Molecular Biology Division, University of California, San Francisco, California 94121, USA

IN mitochondria, electrons derived from the oxidation of succinate by the tricarboxylic acid cycle enzyme succinate–ubiquinone oxidoreductase are transferred directly to the quinone pool. Here we provide evidence that the soluble form of this enzyme (succinate dehydrogenase) behaves as a diode that essentially allows electron flow in one direction only. The gating effect is observed when electrons are exchanged rapidly and directly between fully active succinate dehydrogenase and a graphite electrode. Turnover is therefore measured under conditions of continuously variable electrochemical potential. The otherwise rapid and efficient reduction of fumarate (the reverse reaction) is severely retarded as the driving force (overpotential) is increased. Such behaviour can arise if a rate-limiting chemical step like substrate binding or product release depends on the oxidation state of a redox group on the enzyme. The observation provides, for a biological electron-transport system, a simple demonstration of directionality that is enforced by kinetics as opposed to that which is assumed from thermodynamics.

Succinate dehydrogenase (SDH)¹ consists of two different hydrophilic subunits (F_p and I_p , relative molecular masses 70,000 and 27,000, respectively) containing one covalently bound flavin adenine dinucleotide and three Fe–S clusters (centres 1, 2 and 3). In complex II, succinate–ubiquinone oxidoreductase, these subunits are associated with two hydrophobic peptides that bind the enzyme to the mitochondrial inner membrane. In this physiological form, electrons are exchanged with the quinone pool:



For this study, we used reconstitutively-active SDH (ref. 2), which catalysed succinate oxidation by phenazine methosulphate or low concentrations of ferricyanide³ with k_{cat} (38 °C) = 10–12,000 min^{−1}. An edge-oriented pyrolytic graphite (PGE) electrode was used throughout⁴. We found that cyclic voltammetry of a solution of succinate (pH 8.3 at 35 °C) shows no faradaic activity over the range +200 to −500 mV against a standard hydrogen electrode. But introduction of SDH (0.6–1.2 μM) gives rise to an oxidation current (Fig. 1) which increases rapidly with potential to reach a steady-state value, i_{lim} .

The following observations (1) to (4) are based on first-cycle measurements and show that SDH can adsorb at the pyrolytic graphite electrode in a fully functional electroactive state: (1) i_{lim} is independent of scan rate over the range 2–500 mV s^{−1}. The half-height potential is constant (±10 mV) over the same range, and close to the thermodynamic reduction potential, $E_{F/S}^0$, of the fumarate (F)–succinate (S) couple. Experiments with a rotating-disc electrode⁵ also showed that i_{lim} is independent of rotation rate (0–20 Hz). Thus, neither electron transfer (enzyme to electrode) nor mass transport limit the rate of turnover; (2) i_{lim} is independent of SDH concentration above 0.6 μM. This is consistent with catalysis by enzyme molecules that are saturatively bound at a small fraction of the electrode surface. The adsorbed nature of the catalyst is evident from two further observations: first, that activity increases initially with time at a rate dependent on SDH concentration (for example, below 0.6 μM, i_{lim} increases over the first few cycles at 10 mV s^{−1}); and, second, that electrodes incubated in SDH solution and rinsed with buffer show activity after transfer to a solution without enzyme; (3) i_{lim} varies with succinate

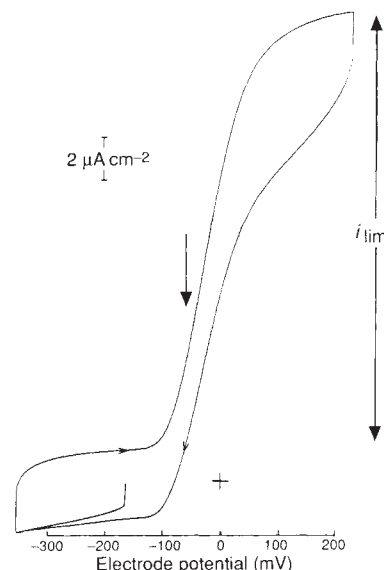


FIG. 1 Cyclic voltammogram at a freshly polished pyrolytic graphite edge electrode of a solution (0.1 M NaCl, 50 mM TAPS (*N*-tris(hydroxymethyl)-methyl-3-aminopropanesulphonic acid)) containing 6.2 mM succinate, 0.7 μM SDH at pH 8.3. Temperature was 35.5 °C, scan rate, 10 mV s^{−1}. The position of $E_{F/S}'$, the formal reduction potential for the fumarate/succinate couple at pH 8.3, is indicated by the vertical arrow. The position of absolute zero current is marked with a cross. The limiting faradaic current i_{lim} is measured as indicated and expressed as current density. Repetitive cycles always showed progressive attenuation of i_{lim} , indicating inactivation of the enzyme. On the other hand, catalytic activity was largely retained if, instead of repetitive cycling, the electrode potential was held for an equivalent period at −200 mV. A likely explanation is degradation of centre 3 at the higher potentials required for succinate oxidation. It is known that loss of low- K_M ferricyanide reductase activity coincides with degradation of centre 3 as SDH is oxidized³. We also found that SDH samples with only 50% and 17% of the ferricyanide reductase activity of pristine enzyme gave respective degrees of diminished electroactivity.

concentration according to Michaelis–Menten kinetics. The K_M value of 0.22 ± 0.02 mM agrees broadly with results from other methods⁶. (4) At pH 8.3 with a saturating concentration of succinate (6.25 mM), the value of i_{lim} yields a rate of succinate oxidation of 4×10^{-9} mol min⁻¹ per square centimetre of electrode surface. From this, k_{cat} is estimated to be 20,000 min⁻¹ on the basis of an upper limit of 5% electrode surface coverage by active SDH (that is, about 2×10^{-13} mol cm⁻², above which we would expect to detect a faradaic response from active sites in the absence of substrate).

Addition of fumarate to the succinate solution produces the striking results shown in Fig. 2. A prominent reduction peak appearing in the direction of decreasing potential is matched by a reduction peak in the direction of increasing potential. The appearance of these 'troughs' is retained in a rotating-disc electrode experiment in which steady-state conditions should be enforced for a diffusion-controlled process⁵. We find, furthermore, that if fumarate is the sole substrate, the pair of troughs is the only faradaic activity observed. We conclude that an increase in the driving force (namely a decrease in applied potential) produces a decrease in the rate of reduction.

This result can be interpreted in terms of a potential-dependent gate that bars catalysis of the reverse process. Thus, while fumarate reduction proceeds rapidly at potentials close to $E_{F/S}^{0'}$ (which we determined as +5 mV at pH 7.5 from variation of the potential of zero net faradaic current with fumarate/succinate ratio), the catalytic activity passes through a maximum (about -85 mV versus $E_{F/S}^{0'}$) and diminishes sharply as the applied potential is lowered. Two further observations lead us to conclude that this is an intrinsic property of the enzyme and not a feature of the electrode–enzyme interaction. First, voltammetry with a polished gold electrode (although the response is much weaker and less stable) showed the trough at exactly the same

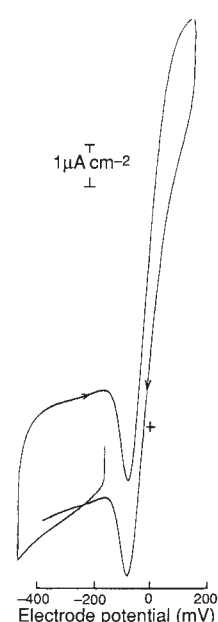


FIG. 2. Cyclic voltammogram at a freshly polished pyrolytic graphite electrode of a solution (0.1 M NaCl, 50 mM HEPES (*N*-(2-hydroxyethyl) piperazine-*N'*-(2-ethanesulphonic acid))) containing 6.2 mM succinate, 1.0 mM fumarate, 1.2 μM SDH at pH 7.5. Temperature was 35.5 °C, the scan rate, 10 mV s⁻¹.

This phenomenon can be accounted for if a rate-determining chemical step is sensitive to the redox status of the catalyst.

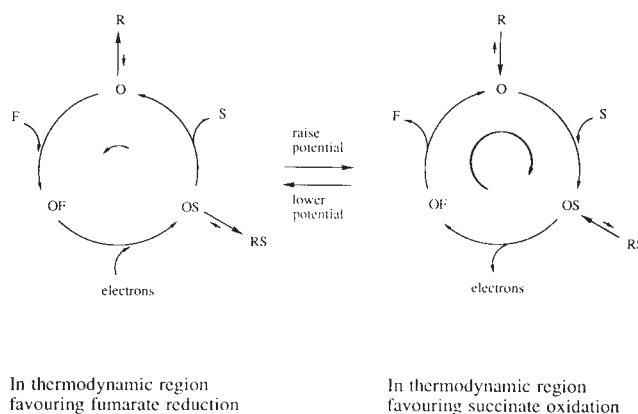


FIG. 3 Scheme illustrating how SDH catalysis of the fumarate(F)/succinate(S) system can be gated by an excessive driving force which elevates the steady-state level of R (FADH₂), a form of the active site in which substrate entry or product release does not occur readily.

Electrochemical behaviour of this type is found in relatively simple catalytic systems⁸. Our observation with SDH may be rationalized if substrate binding or product release are allowed only during the period when the active site is oxidized (that is, in the FAD form). This restriction is indeed suggested from steady-state kinetic studies⁹. The resulting effect is visualized in Fig. 3. For simplification, only oxidized 'O' (FAD) or reduced 'R' (FADH₂) forms of the active site are considered, and the participation of H⁺ is omitted. The enzyme–substrate/product intermediate 'RF' is not included. It is assumed to exist only transiently, either because equilibrium favours the isoelectronic species 'OS' (spontaneous electron redistribution at the active site), or because spontaneous electron transfer to the Fe–S centres and electrode rapidly produces 'OF'. In the electrochemical experiment, the direction and rate of electron transfer are continuously variable through the applied electrode potential. As this is lowered (thermodynamically favouring fumarate reduction), the lifetime (and population) of 'O' is decreased; conditions instead favour the formation of 'R' and catalysis is blocked. The effect is analogous to inhibition by a substrate (in this case, the electron). By contrast, as the applied potential is raised (thermodynamically favouring succinate oxidation), the lifetime (and population) of 'O' is increased, and the reaction proceeds smoothly in the physiologically observed direction. Catalysed fumarate reduction is thus restricted to a narrow window of potential.

Electrochemically, SDH behaves here in a manner resembling the current–voltage characteristics of a tunnel diode¹⁰. The capability is thus demonstrated for an electron-transport enzyme to 'rectify' electron flow at potentials close to electrochemical reversibility. Such a feature may be used *in vitro* for regulation. The observation is attributable to the ability, afforded by the direct voltammetric method, to measure the turnover rate in response to a continuously variable driving force. □

Received 16 September 1991; accepted 7 January 1992.

- Ackrell, B. A. C., Johnson, M. K., Gunsalus, R. P. & Cecchini, G. in *Chemistry and Biochemistry of Flavoenzymes* Vol. 3 (ed. Mueller, F.) 229–297 (CRC, Boca Raton, Florida, 1992).
- Davis, K. A. & Hatefi, Y. *Biochemistry* **10**, 2509–2516 (1971).
- Beinert, H., Ackrell, B. A. C., Vinogradov, A. D., Kearney, E. B. & Singer, T. P. *Archs Biochem. Biophys.* **182**, 95–106 (1977).
- Armstrong, F. A. *Struct. Bond.* **72**, 137–221 (1990).
- Bard, A. J. & Faulkner, L. R. *Electrochemical Methods* Ch. 8 (Wiley, New York, 1980).
- Singer, T. P., Kearney, E. B. & Kenney, W. C. *Adv. Enzym.* **37**, 189–272 (1973).
- Ohnishi, T. et al. *J. Biol. Chem.* **256**, 5577–5582 (1981).
- Collman, J. P. & Kim, K. J. *Am. chem. Soc.* **108**, 7847–7849 (1986).
- Zeijlmeijer, W. P., Dervartanian, D. V., Veeger, C. & Slater, E. C. *Biochim. biophys. Acta* **178**, 213–224 (1968).
- Mackenroth, D. R. & Sands, L. G. *Illustrated Encyclopaedia of Solid-state Circuits and Applications* Ch. 1 (Prentice-Hall, Englewood Cliffs, New Jersey, 1984).

Rewards of working for the government

Once-secure employment in the public service is now threatened by the brooding of most governments about its cost, but the rewards (intellectual and otherwise) may still be substantial.

WORKING for the government, whichever government it may be, is not usually regarded as a glamorous occupation for researchers, but that does not necessarily mean that it is unrewarding, financially or even intellectually. And sometimes there is no alternative. In the Socialist countries of Central Europe and in the ex-Soviet Union, governments were constitutionally the sole employers of scientifically trained people, and are likely to remain so for some time. In China, even though the process of economic reform seems now to have been resumed, it will be some time before enterprises large enough to employ substantial numbers on research and development will become economically independent. It would be surprising if, worldwide, the proportion of technical people working for the governments of one stripe or another were less than a half.

Even in Europe and North America, the public service is glutinous in its consumption of technical skill. In Britain, for example, the Institution of Professional and Managerial Staffs estimates that there are 19,900 people with scientific and technical qualifications working for central government in various ways. But the tentacles of the public service reach beyond government offices and laboratories, often far beyond them.

Academics

In Italy and France, for example, university teachers are technically employees of their central governments, which determine salaries, conditions of employment and overall numbers of those employed. (In Germany, the *Länder* play that role.) But even where universities are titularly autonomous, as in Britain or The Netherlands, central government

eventually foots the bill for teachers' and researchers' salaries, so that academics and their representatives find themselves negotiating with civil servants on pay and conditions of service.

Among major countries, only in the United States, Japan and Switzerland is the influence of central government conspicuously less obtrusive. The explanation is two-fold. First, the proportion of total research and development funds spent by companies and other commercially autonomous organizations is at the upper end of the spectrum. Second, in all three countries (all technically advanced, not necessarily by coincidence) arrangements have been made whereby technical people engaged on central government's business are formally employed by autonomous organizations standing at least an arm's length away from government.

Public service

The now-general suspicion of the public service flies in the face of much tradition over the past century or so. Who but the US Army Corps of Engineers built Boulder Dam? In the same vein, the US Army Medical Corps identified the vector of yellow fever; Watson-Watt worked out a practicable scheme for the use of radar in the defence of Britain against air attack. The public service used to be an heroic service. Why has its reputation so much changed?

Part of the explanation is the commonplace observation that science at large has grown more quickly than government science. That is everywhere the case. Although people may argue about comparisons of the attention (and funds, as a proportion of Gross Domestic Product) paid by different governments in support of

untied research, the truth is that all central governments now spend a smaller proportion of their countries' total spending on research in-house than, say, half a century ago.

Another component of an explanation is that public-service laboratories and other publicly-managed technical organizations have become intellectually timorous, partly because they are now less sure of themselves than they used to be. A government laboratory bent on demonstrating, for example, that the global warming is already (as it may be) a reality would have to run the gauntlets not only of his or her sceptical untrained civil servant colleagues, but of much more expert sceptics in the academic world. Who could blame a beginning civil servant, in circumstances like that, for deciding to urge his civil servant masters to let a contract to some university in the hope that his or her own argument will be echoed?

There are also conditions of employment to be considered. Public servants, by definition, are responsible to the government. (In Britain, they are usually also bound by the Official Secrets Act which, even as recently amended, onerously inhibits a person's free discussion, as over coffee with a colleague, of what he is about.) Although the consequences of these restraints are more often demeaning than impediments to useful work, they weigh with people. The public service remains a secret service in an awkward sense.

A further difficulty is the general uniformity of conditions of employment. Public servants differ from ordinary mortals in their expectation that salaries, for example, will increase with the numbers entered in a matrix whose axes represent age (or

years of service) and some measure of qualifications. Many spouses (and bank managers) have been comforted by those built-in expectations.

By contrast, public servants are relatively much more constrained than colleagues in other settings by the rules with which they must comply: allowable work-related expenses are rigorously defined, even hours of work may be restricted by arguments about laboratory safety while an impulse to make a public comment, even about an issue unrelated to one's work, may have to be negotiated with an unsympathetic director.

Private influence

The other side of that coin is easily described: public-servant scientists and researchers can hope to trade lost freedom for a measure of power and influence beyond the ken of academic scientists. There is little doubt that such a consideration (job-security apart) explains why so many people worldwide are eager for government appointments. And nobody disputes that able people with strong opinions on issues of public importance can powerfully influence policy from even novitiate positions within the public service.

Most civil services are perennially short of technical people able to speak intelligently, or even to speak, at committees where the issues raised may be technical in character. These are great opportunities for those who can think on their feet, and who are ready for the risks of being outsmarted by undisclosed expertise elsewhere around the table.

Surprisingly, the public service everywhere remains a haven for scholarship of an important kind — of a kind of which it knows it has too little. It is obvious that the accumulation of experience of practical problems in fields such as the hazards to fish of toxic materials in river water, the occurrence of novel diseases among domestic animals (witness British excitement three years ago about the incidence of bovine spongiform encephalitis, or BSE) and in the assessment of the dangers of

particular nuclear accidents, is a powerful spur to the evolution of sagacity in the public service.

It is less obvious, but no less true, that the public service is, even now, well-equipped to assist with the solution of important practical problems. In the United States, most of the national laboratories are busy on the exploitation of "alternative" (and renewable) forms of energy. The laboratories of the US Department of Agriculture are slowly girding their loins for the genetic engineering of plants and animals — and for an exploration of the difficulties. At half a dozen British public laboratories, there are teams of able people working hard and interestingly on a national inventory of the sources of greenhouse gases (which is far from child's play).

In short, the public service, almost like a university, is much what those who enter it choose to make of it. To be sure, entrants should in prudence take care that they do not enter the public service blind, without knowing where they will be working and what they will be doing. It is even prudent that they should discover in advance whether their intended unit or laboratory is considered to be successful or the opposite. There is nothing worse, at the beginning of a career, than being part of the terminal decline of an institution. Even the certainty of job-security cannot make up for that.

Rewards

That is especially so because the public service everywhere is relatively mean with money. That requires no special explanation. The managers say that security (and often a generous and index-linked pension) should be a sufficient recompense for that. But the doctrine that public service is a public good and therefore an estimable public service still carries weight.

The immediate result, for example, is that entrants to the public service are usually paid less well than those who take up work for commercial organizations, sometimes in rela-

tively menial roles (as technicians rather than researchers in the strict sense). In Britain now, for example, first graduates from British universities can expect a starting salary of £11,000, plus or minus 10 per cent depending on the rarity of their speciality and the proposed location of their work. Organizations providing commercial financial services would probably offer similar people at least half as much again. Most other commercial organizations would come somewhere in between.

For entrant public servants, however, there is an important difference: salaries are determined by length of service as well as by quality. Japan, as always, is also different. Traditionally, those who worked in the sprinkling of government laboratories were no better equipped (or paid) than simple academics. The past five years have seen important changes. Although the favoured route into research depends on the patronage of a professor and the interest of a company, the public laboratories are no longer housed in buildings run up in the aftermath of the 1923 earthquake. With luck and a modicum of flair, a person can hope to win a reputation as a scientist as easily as one who happens to work at Los Alamos or Oak Ridge. Japan's public service is much more attractive than it used to be, if the technical people who really count are those who call the shots at the ministries.

The public service everywhere is thus too easily undervalued as a career. It is one of those occupations in which the disadvantages are more easily catalogued than the rewards. Of course, there are dangers, those of being cut off from science by the bureaucracy and those of becoming a bureaucrat. (A British civil servant asked for information about public servants' pay declined to provide it for fear that it might cause political trouble, especially during an election campaign.) The trick is to decide what you want before you join up and then to stick to your resolve; the chances are that the public service will give you the opportunity. □